A *Carbon Data Model framework* to generate probabilistic estimates of ecosystem traits and processes

Mathew Williams, Luke Smallman + Global Change Ecology Lab

University of Edinburgh and NCEO, Edinburgh, UK

Digital Twins for the Terrestrial Carbon Cycle
ESA Carbon from Space Workshop

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

What is a digital twin and why is it useful?

A computer program

that takes as inputs

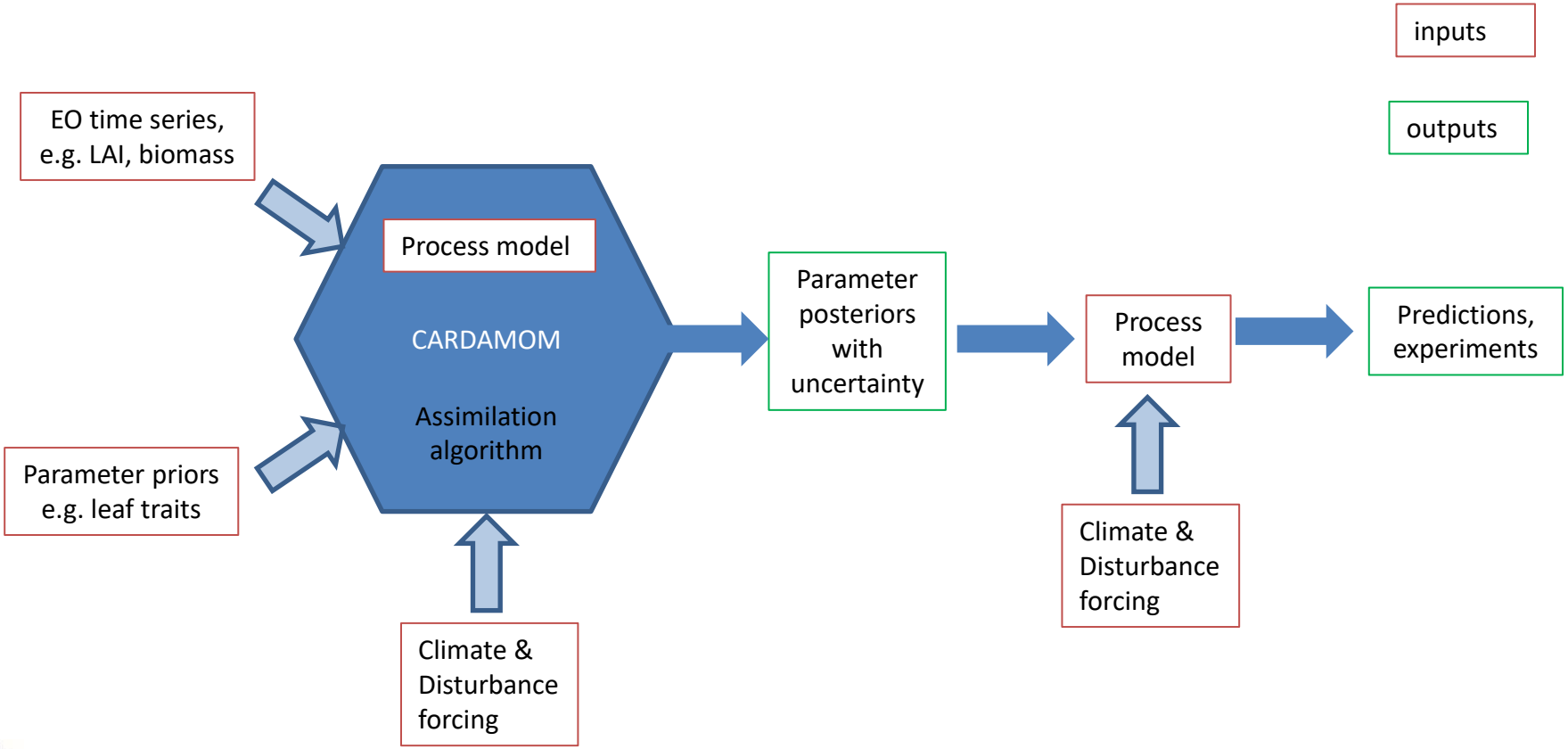
->real-world data about a physical system

and produces as outputs

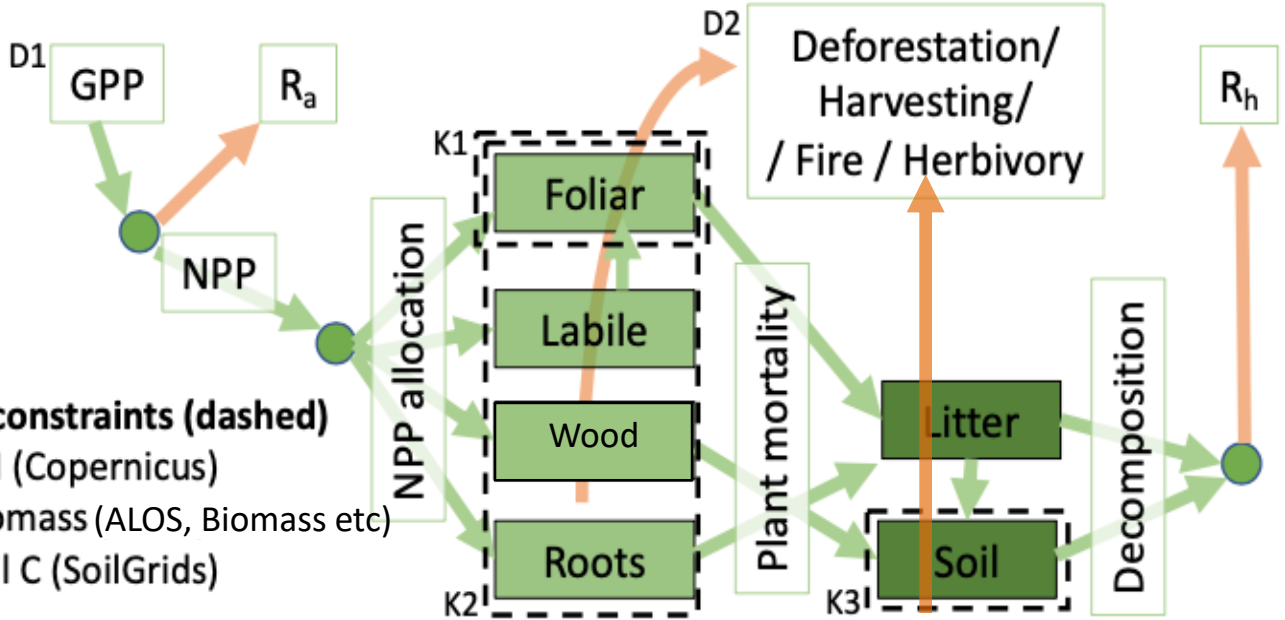
->predictions or simulations of how that physical system will be affected by those inputs

DT can inform global C models, identify sources of forecast error, quantify the information content of EO data, & directly link land use data to C assessments

A C Digital Twin Approach



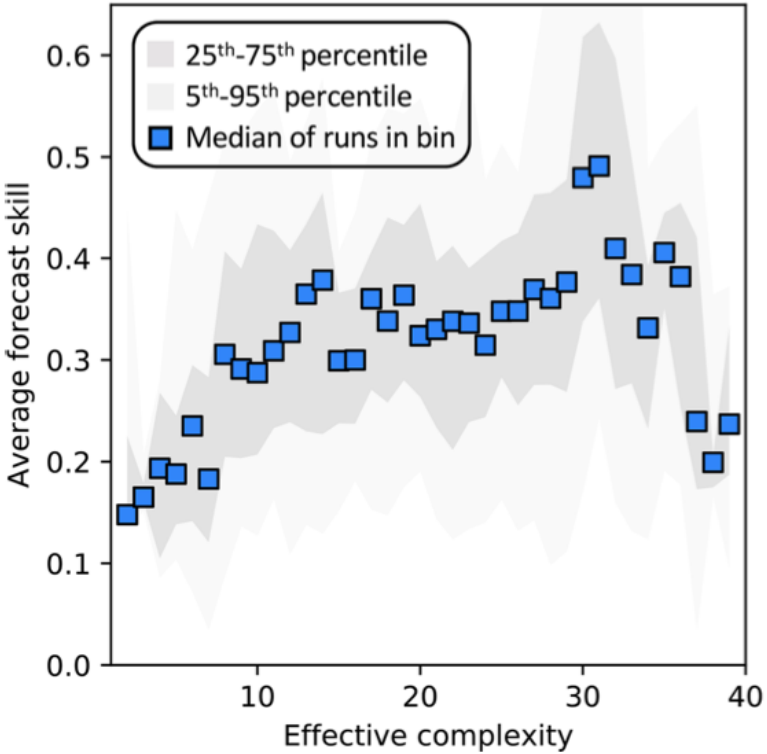
Global calibration via EO data links

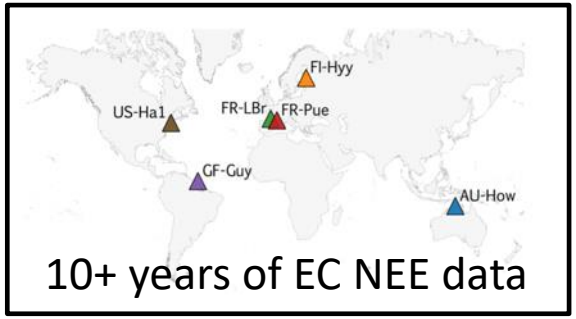


Data constraints (dashed)
K1 LAI (Copernicus)
K2 Biomass (ALOS, Biomass etc)
K3 Soil C (SoilGrids)

Model drivers
D1 Meteorology (various)
D2 Deforestation, tree harvesting (GFW) Fire (MODIS), Herbivory (FAO/Hempson et al)

Complexity ≠ forecasting skill





10+ years of EC NEE data

Famiglietti et al 2021

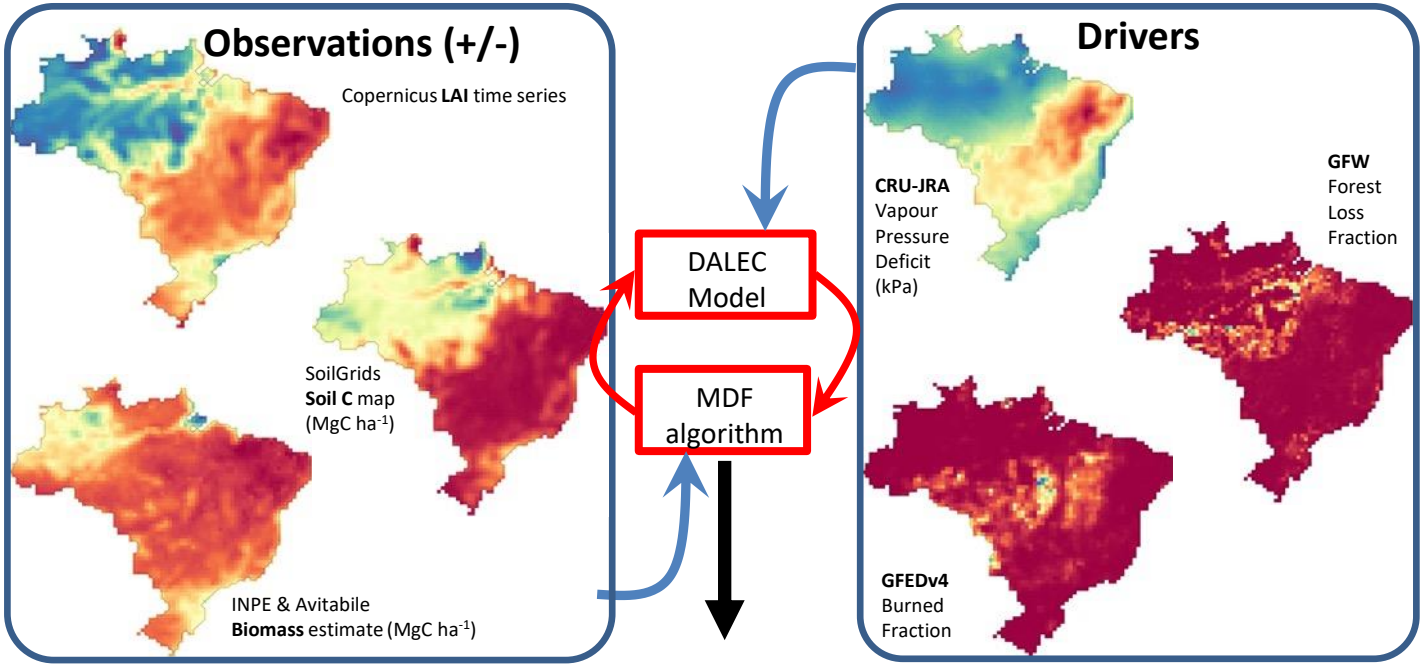
50% of EC NEE data retained for forecast validation

DT: importance of model structure, parameter constraint, forcing effects on forecasts

Model	Photosynthesis	Water cycle	Plant respiration	Wood litter	No. of parameters
M1	ACM1	No	R_a : GPP	No	23
M2	ACM2	No	R_a : GPP	No	23
M3	ACM2	Yes	R_a : GPP	No	23
M4	ACM2	Yes	R_m : GPP + R_g : NPP	No	27
M5	ACM2	Yes	R_m : GPP + R_g : NPP	Yes	29

Smallman et al. 2021

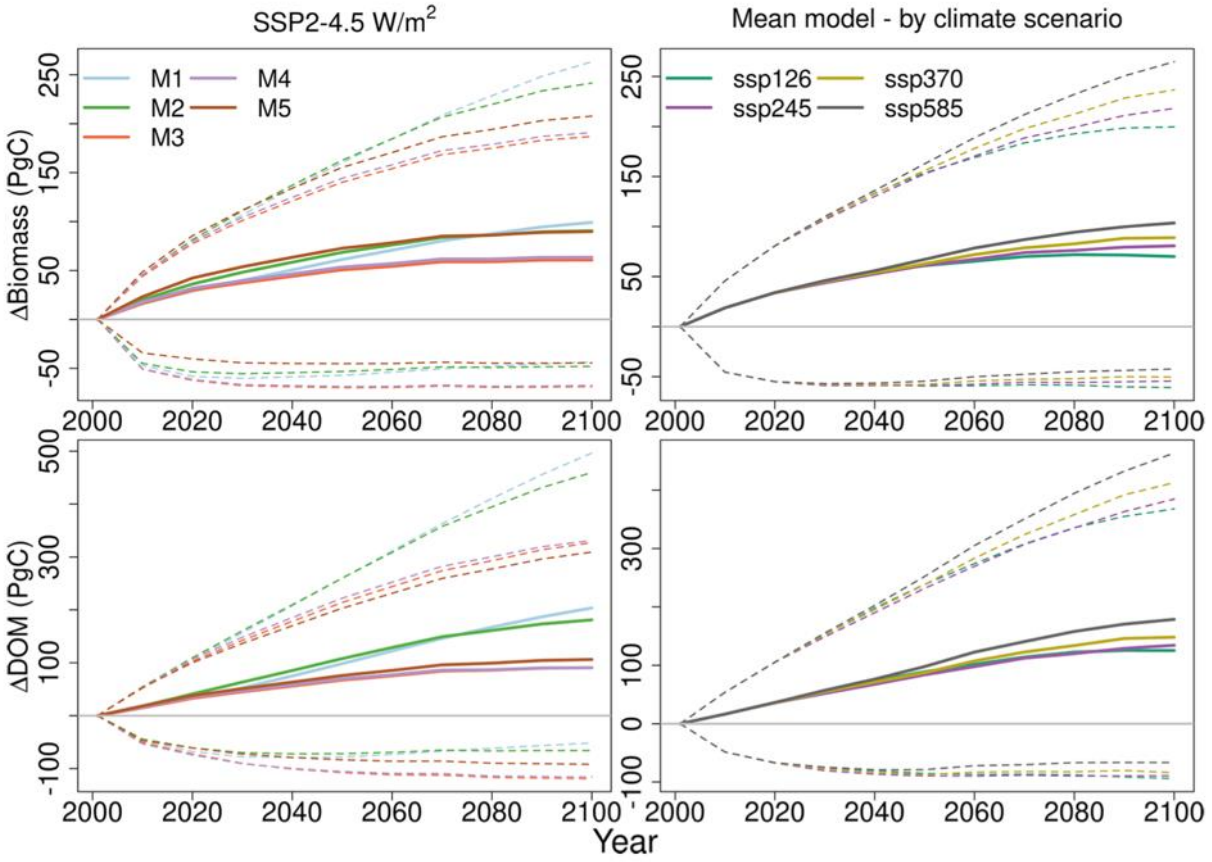
CARDAMOM Assimilation Scheme



Output: Probabilistic estimates of parameters, C pools and fluxes for each pixel

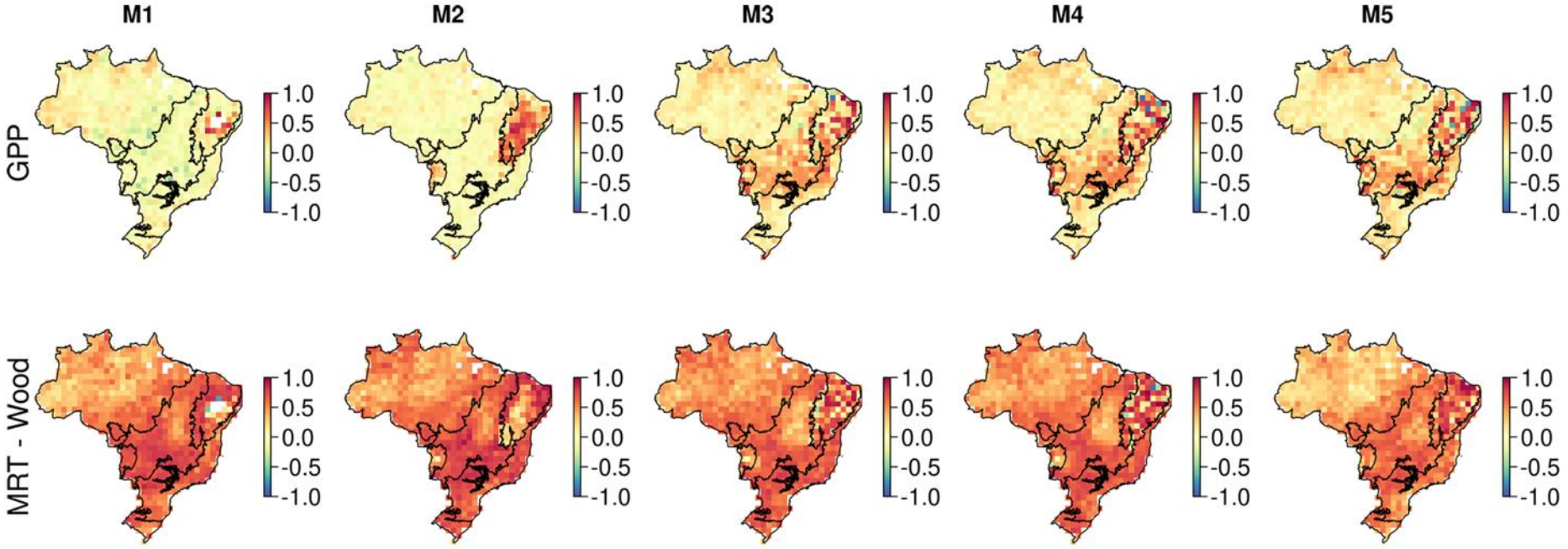
Forecast: C pools & fluxes to 2100 under varied scenarios

Parameter uncertainty dominates forecast spread

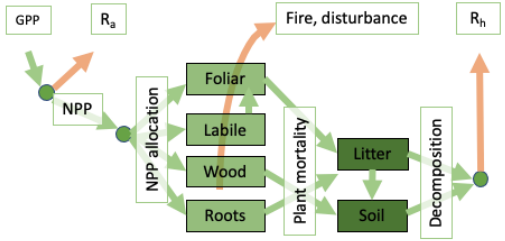


Smallman et al. 2021

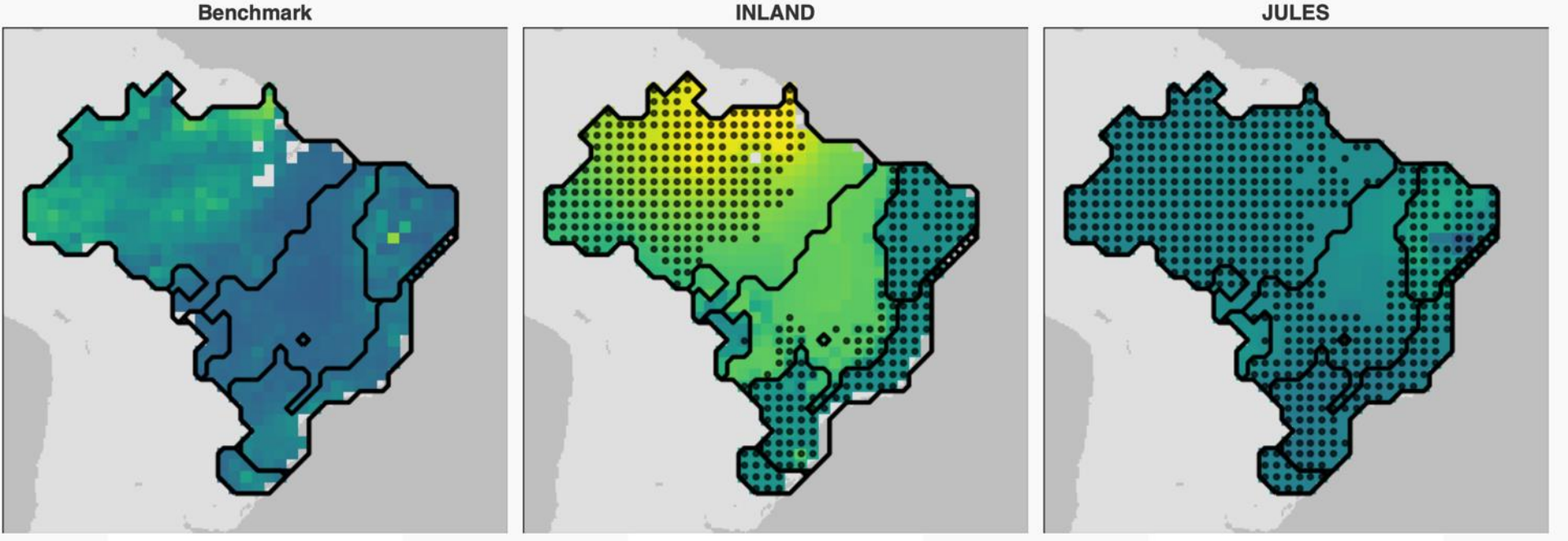
Correlation maps between the simulated change in biomass stocks (2001–2100) and ecosystem variables



(1) MRT_{wood} and (2) NPP_{wood} are the parameters most tightly coupled to the response of biomass C stocks to climate change



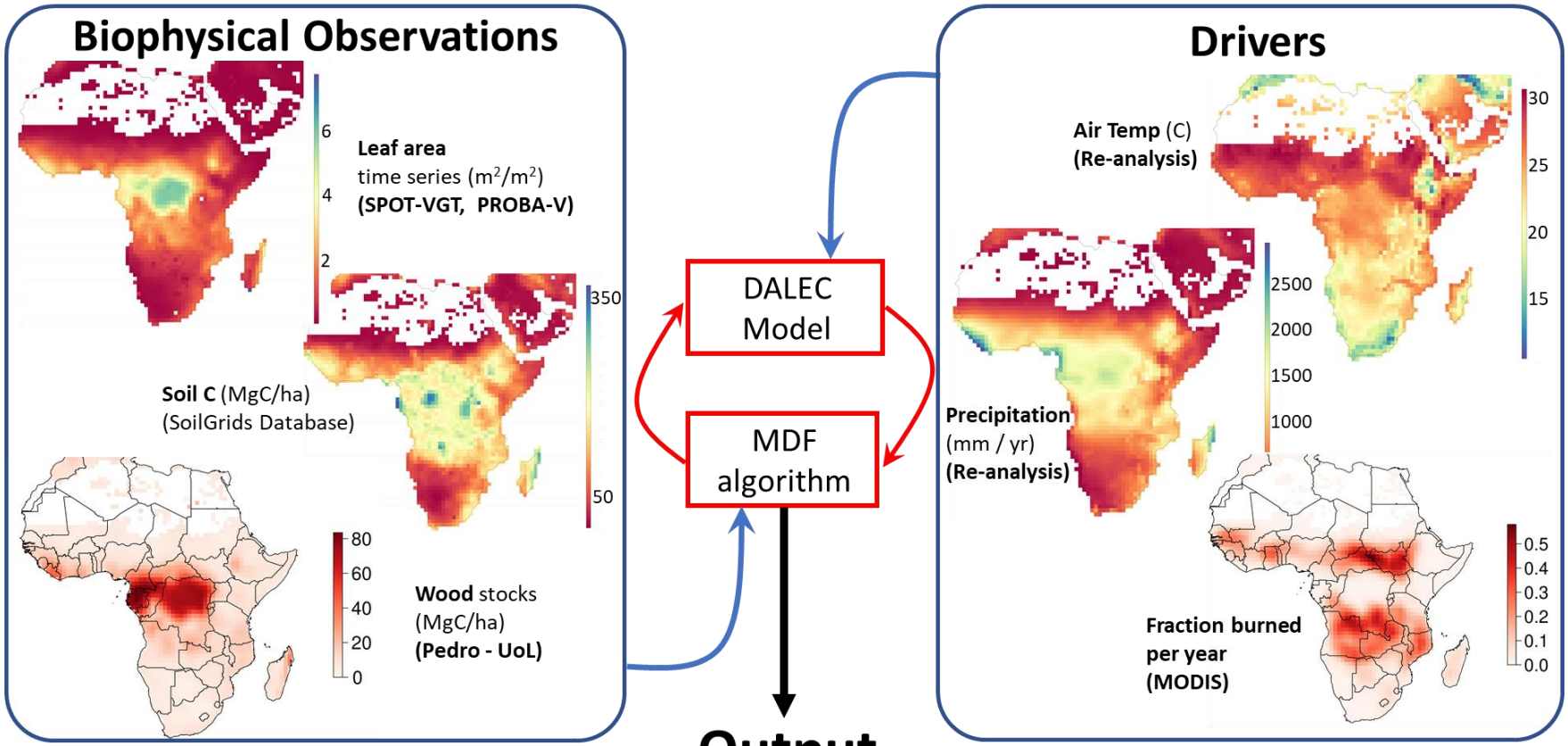
Contemporary wood residence time: CARDAMOM vs two Forecast Models



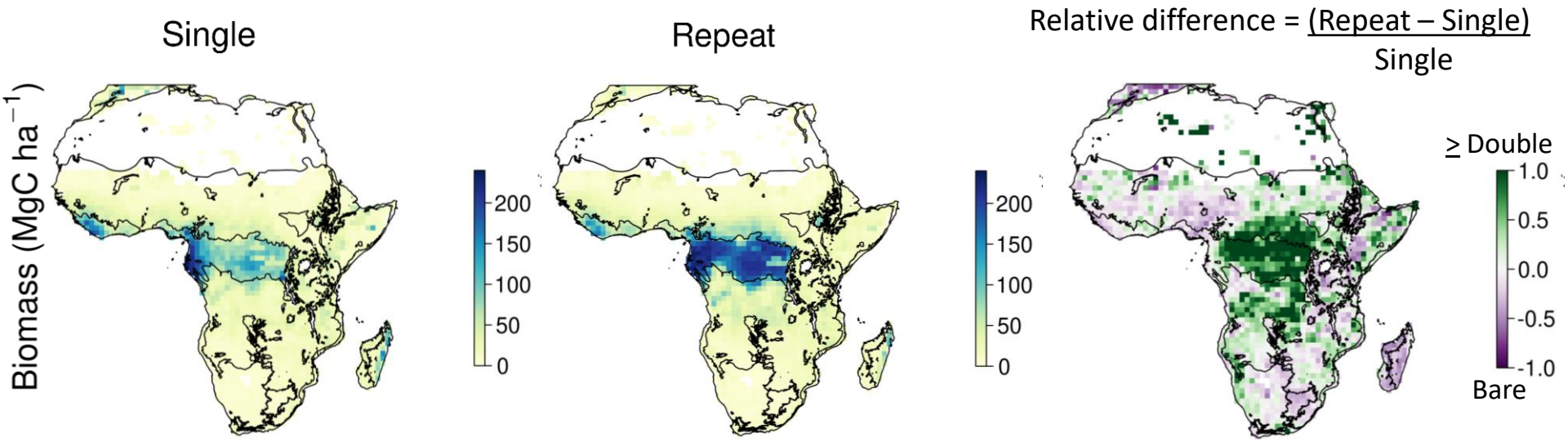
Caen et al. 2021

Stipples show LSM is within 90% CI

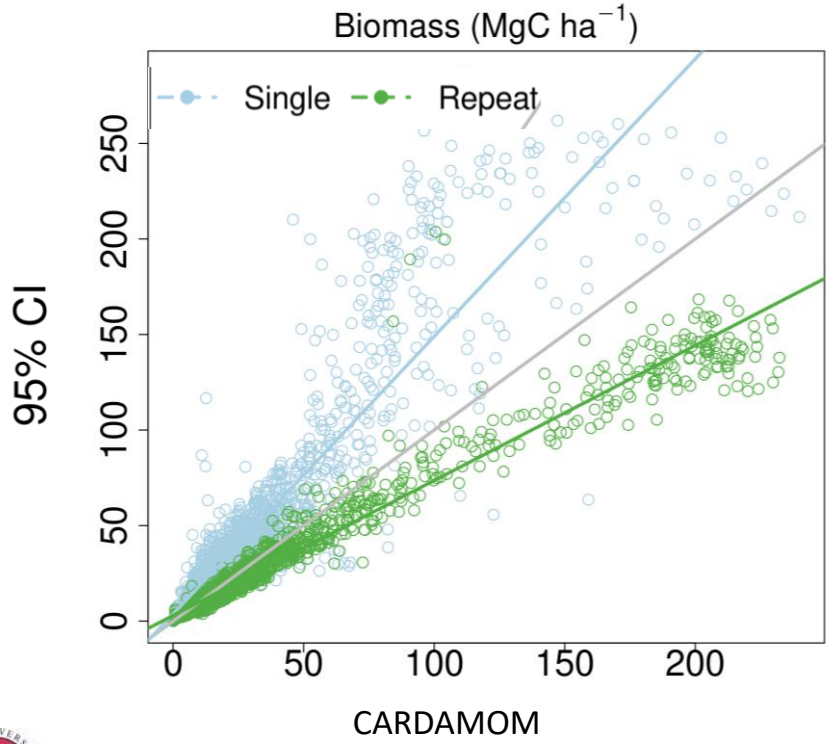
DT to evaluate information content of EO



Estimated biomass is strongly impacted

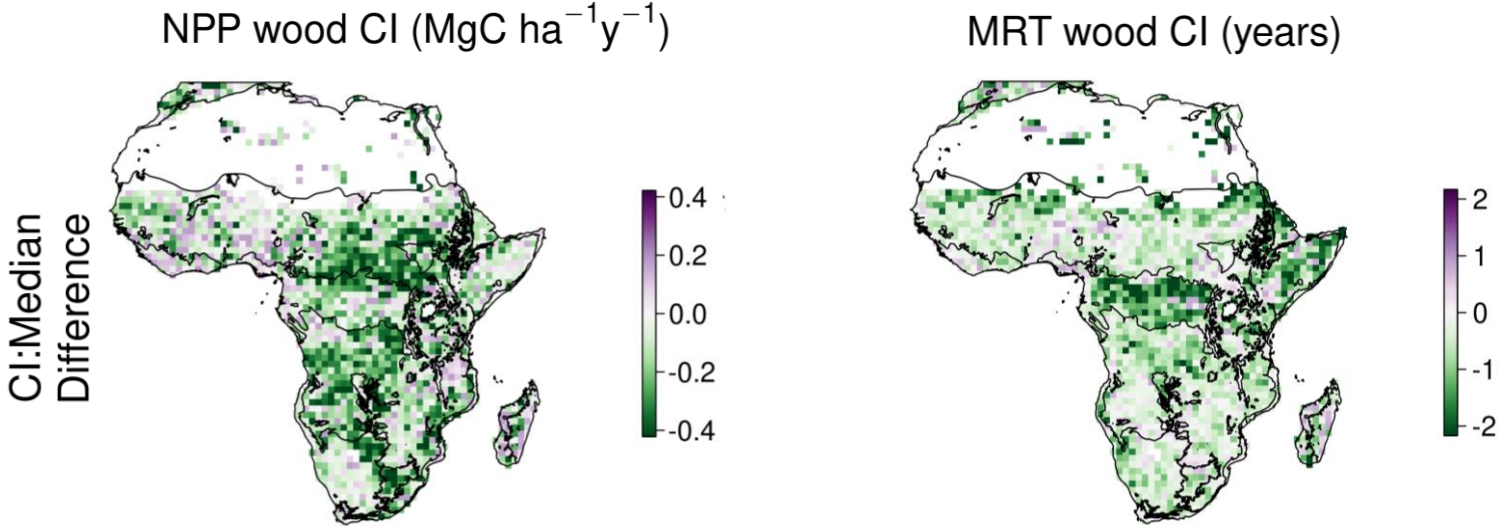


Biomass uncertainty reduced



$$\text{CI:Median} = \frac{\text{CI}}{\text{Median}}$$

Process constraint:
Residence time and NPP uncertainty reduced



$$\text{CI:Median} = \frac{\text{CI}}{\text{Median}}$$

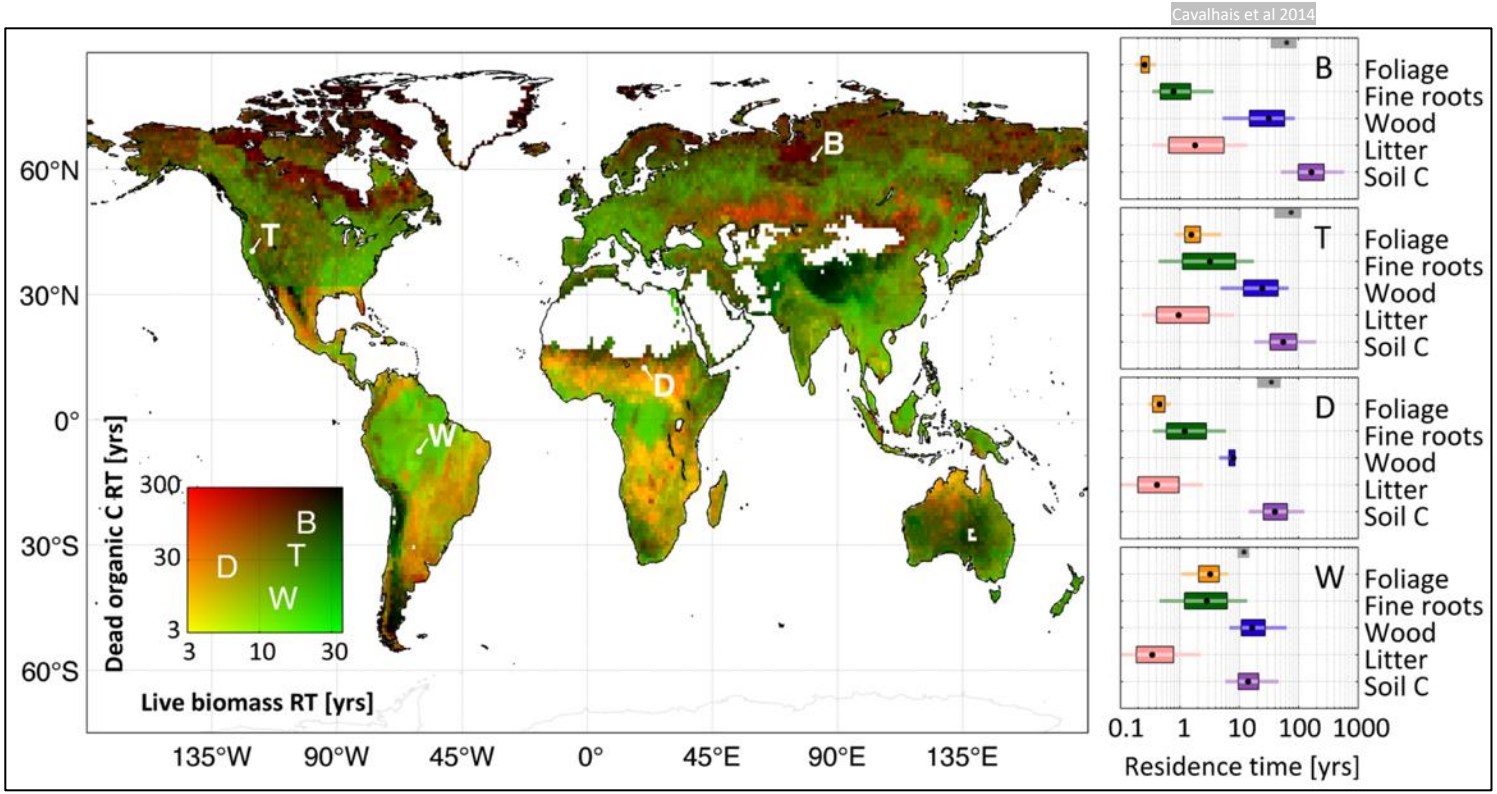
Summary – Digital Twins can

- Identify sources of forecast error, *e.g. from drivers, parameters, structures to focus next steps*
- Inform global C models, *e.g. evaluate process representation, like mean residence times*
- Quantify the information content of EO data, *e.g. determine the value of repeat observations for process constraint*
- Support net zero counter-factuals via probabilistic model experiments; e.g. alternate land use, management (ref: land use talk from Wednesday)

Acknowledgements:
Luke Smallman, Jeff Exbrayat, Anthony Bloom, David Milodowski, Pedro Rodriguez Viega, Caroline Famiglietta, Alex Konings
CARDAMOM Club, INPE, GEDI, ALOS
JULES and INLAND team

Funding: NERC, NCEO, Newton Fund, UKSA

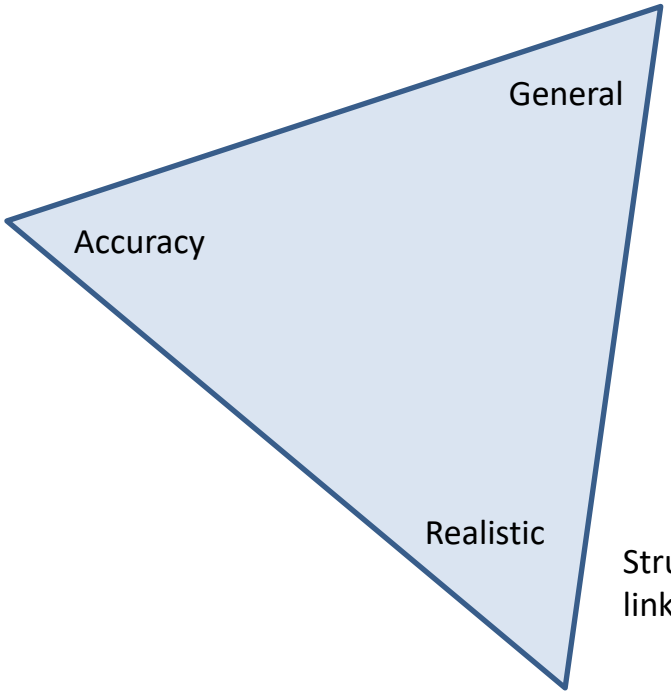
Global retrievals of carbon residence times



Residence time ~ Pool lifespan

Bloom et al., PNAS 2016

Model Trade-offs



Lean requirements
for local-global application

Calibrated
Avoid over-fitting
Validated
Uncertainties

Structure represents key processes,
linked to model complexity

Accurate and general models suffice