AGU ADVANCING EARTH AND SPACE SCIENCE

Reviews of Geophysics

REVIEW ARTICLE
10.1029/2020RG000715

# Reanalysis in Earth System Science: Toward Terrestrial Ecosystem Reanalysis

**R. Baatz**[1,2], **H. J. Hendricks Franssen**[1], **E. Euskirchen**[3], **D. Sihi**[4], **M. Dietze**[5], **S. Ciavatta**[6,7], **K. Fennel**[8], **H. Beck**[9], **G. De Lannoy**[10], **V. R. N. Pauwels**[11], **A. Raiho**[12], **C. Montzka**[1], **M. Williams**[13], **U. Mishra**[14], **C. Poppe**[1], **S. Zacharias**[15], **A. Lausch**[16,17], **L. Samaniego**[18], **K. Van Looy**[19], **H. Bogena**[1], **M. Adamescu**[20], **M. Mirtl**[15], **A. Fox**[21], **K. Goergen**[1,22], **B. S. Naz**[1,22], **Y. Zeng**[23], and **H. Vereecken**[1,2]

[1]Agrosphere, Institute of Bio and Geosciences, Forschungszentrum Jülich, Jülich, Germany, [2]Scientific Coordination Office International Soil Modelling Consortium ISMC, Jülich, Germany, [3]University of Alaska Fairbanks, Institute of Arctic Biology, Fairbanks, AK, USA, [4]Department of Environmental Sciences, Emory University, Atlanta, GA, USA, [5]Earth and Environment, Boston University, Boston, MA, USA, [6]Plymouth Marine Laboratory, Plymouth, UK, [7]National Centre for Earth Observation, Plymouth Marine Laboratory, Plymouth, UK, [8]Department of Oceanography, Dalhousie University, Halifax, NS, Canada, [9]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA, [10]Department of Earth and Environmental Sciences, KU Leuven, Heverlee, Belgium, [11]Department of Civil Engineering, Monash University, Clayton, VIC, Australia, [12]Fish, Wildlife, and Conservation Department, Colorado State University, Fort Collins, CO, USA, [13]School of GeoSciences and NCEO, University of Edinburgh, Edinburgh, UK, [14]Bioscience Division, Sandia National Laboratory, Livermore, CA, USA, [15]Department of Monitoring and Exploration Technologies, UFZ Helmholtz Centre for Environmental Research, Leipzig, Germany, [16]Department Computational Landscape Ecology, Helmholtz Centre for Environmental Research–UFZ, Leipzig, Germany, [17]Geography Department, Humboldt University Berlin, Berlin, Germany, [18]Department Computational Hydrosystems, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany, [19]OVAM, International Policy Unit, Mechelen, Belgium, [20]Research Center for Systems Ecology and Sustainability, University of Bucharest, Bucharest, Romania, [21]Joint Center for Satellite Data Assimilation, UCAR, Boulder, CO, USA, [22]Centre for High-Performance Scientific Computing in Terrestrial Systems, Geoverbund ABC/J, Jülich, Germany, [23]Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands

**Key Points:**

- Reanalyzes provide decades-long model-data-driven harmonized and continuous data sets for new scientific discoveries
- Novel global scale reanalyzes quantify the biogeochemical ocean cycle, terrestrial carbon cycle, land surface, and hydrologic processes
- New observation technology and modeling capabilities allow in the near future production of advanced terrestrial ecosystem reanalysis

**Correspondence to:**
R. Baatz,
r.baatz@fz-juelich.de

**Citation:**
Baatz, R., Hendricks Franssen, H. J., Euskirchen, E., Sihi, D., Dietze, M., Ciavatta, S., et al. (2021). Reanalysis in Earth system science: Toward terrestrial ecosystem reanalysis. *Reviews of Geophysics*, *59*, e2020RG000715. https://doi.org/10.1029/2020RG000715

Received 12 OCT 2020
Accepted 20 JUN 2021

**Abstract** A reanalysis is a physically consistent set of optimally merged simulated model states and historical observational data, using data assimilation. High computational costs for modeled processes and assimilation algorithms has led to Earth system specific reanalysis products for the atmosphere, the ocean and the land separately. Recent developments include the advanced uncertainty quantification and the generation of biogeochemical reanalysis for land and ocean. Here, we review atmospheric and oceanic reanalyzes, and more in detail biogeochemical ocean and terrestrial reanalyzes. In particular, we identify land surface, hydrologic and carbon cycle reanalyzes which are nowadays produced in targeted projects for very specific purposes. Although a future joint reanalysis of land surface, hydrologic, and carbon processes represents an analysis of important ecosystem variables, biotic ecosystem variables are assimilated only to a very limited extent. Continuous data sets of ecosystem variables are needed to explore biotic-abiotic interactions and the response of ecosystems to global change. Based on the review of existing achievements, we identify five major steps required to develop terrestrial ecosystem reanalysis to deliver continuous data streams on ecosystem dynamics.

**Plain Language Summary** A reanalysis is a unique set of continuous variables produced by optimally merging a numerical model and observed data. The data are merged with the model using available uncertainty estimates to generate the best possible estimate of the target variables. The framework for generating a reanalysis consists of the model, the data, and the model-data-fusion algorithm. The very specific requirements of reanalysis frameworks have led to the development of Earth-compartment specific reanalysis for the atmosphere, the ocean and land. Here, we review atmospheric and oceanic reanalyzes, and in more detail biogeochemical ocean and terrestrial reanalyzes. In particular, we identify land surface, hydrologic, and carbon cycle reanalyzes which are nowadays produced in targeted projects for very specific purposes. Based on a review of existing achievements, we identify five major steps required to develop reanalysis for terrestrial ecosystem to shed more light on biotic and abiotic

interactions. In the future, terrestrial ecosystem reanalysis will deliver continuous data streams on the state and the development of terrestrial ecosystems.

## 1. Introduction

A reanalysis of a component of the Earth system provides a physically, chemically, and biologically consistent description of continuous past model states by merging multi-source observations and computational models (e.g., Balsamo et al., 2015; Bosilovich et al., 2008; Hersbach et al., 2020; Kalnay et al., 1996; Lorenz & Kunstmann, 2012; Saha et al., 2010). Reanalyzes in Earth system science provide extremely valuable data sources on the past states of (components) of the Earth system, like the atmosphere and ocean. These reanalyzes can serve as input for other model calculations (e.g., land surface models forced by atmospheric reanalysis data), lead to increased process understanding, provide data for states and fluxes which cannot be directly observed and are also valuable data sources for industry and policy. While atmospheric and ocean reanalysis have a tradition dating back to the 1970s (e.g., Bengtsson et al., 1982; "Planning for First Garp Global Experiment (Fgge)," 1972), reanalyzes for terrestrial processes such as land-atmosphere interaction (Section 2.3), hydrology (Section 2.4), and terrestrial carbon cycling (Section 2.5) has only recently gained traction. The most recent developments in biogeochemical reanalysis are of particular importance for ecosystem science as these reanalyzes focus more on life-supporting biotic processes at the Earth's surface.

Knowledge on ecosystem change, management of ecosystems and ecosystem functions are important to society as ecosystems provide invaluable services to humanity such as denaturing pollutants, improving water quality, food production, resistance to pest and disease, mitigating climate change and enhancing soil resilience (Costanza et al., 1997; Lal, 2015). Natural and managed ecosystems are complex, structured systems composed of living (biotic) and non-living (abiotic) components. Abrupt climate change and human induced land use changes are inducing strong pressures on the environment and are diminishing ecosystem values and services (Kubiszewski et al., 2017; Vitousek et al., 1997). Acknowledging and estimating the value of ecosystem services to society and its sustainable development goals (SDGs) are important (Wood et al., 2018) and an essential prerequisite to counteracting negative impacts of anthropogenic activities, leading to more sustained ecosystem management and application of nature based solutions (Guerry et al., 2015; Motesharrei et al., 2016; Rosa et al., 2020). Forecasting the Earth and ecosystem behavior under global change requires an understanding of the key processes controlling ecosystem changes, which is subsequently indispensable for designing and implementing measures and actions to meet the SDGs. A clear analysis of past and current states and dynamics of ecosystems are also critical to forecasting future scenarios (Bonan & Doney, 2018; Rosa et al., 2020).

Terrestrial ecosystem reanalysis aims to provide this analysis of past and current states of the ecosystem. A terrestrial ecosystem reanalysis is a continuous set of states and fluxes (and possibly parameters) of the interacting biophysical compartments of the ecosystems, holistic, and coherent to the model used and including mechanistic process representation. It provides a model-data-driven initialization for the assessment of the future ecosystem development under scenarios of global change. The system descriptions underlying the existing land surface, carbon and hydrologic cycles models also provide a basis for developing a conceptual description of a terrestrial system reanalysis by linking the individual models. Terrestrial ecosystem reanalysis goes beyond current state-of-the-art land surface, hydrologic, and carbon cycle reanalysis by also turning biotic ecosystem observations and Earth observation data into continuous ecologically meaningful maps. An ecosystem reanalysis bridges the gap between remote sensing products, *in situ* observation networks common in biodiversity research and a terrestrial ecosystem model with different availabilities and quality of observation data over the historical period considered. Examples of such in situ observations suitable for assimilation are above- and below-ground biomass, crop yield, net ecosystem exchange, and plant traits composition, and examples of remotely sensed observations are leaf area index, canopy height, and aboveground biomass, gross primary production and plant phenology. We distinguish model-data fusion-based reanalyzes from data-driven reconstructions that use correlation, interpolation and analysis techniques to create best estimates. In contrast to reanalysis, such data-driven reconstructions do not fuse the data with a model along the simulation period.

Data assimilation methods or model-data fusion methods optimally combine Earth system models and observations in a Bayesian sense. They constitute the methodologic basis for reanalysis of the Earth system provided with datasets and products consistent with the underlying physical, chemical and biological principles (Bennett & Budgell, 1987; Evensen, 2003; Kalnay, 2003). Uncertainty in reanalysis products is explicitly considered depending on the data assimilation method. For example, with an ensemble-based data assimilation approach, different model trajectories are calculated which are all consistent with past measurement data. Uncertainty estimates can be derived from the ensemble statistical measures of the model (e.g., variances, covariances). Model-data fusion methods are also an essential ingredient of now- and forecasting systems of which ecological forecasting has gained significant traction in recent years (Dietze, 2017). Adjoined to ecosystem reanalysis, ecological forecasting is defined to be the prediction of ecosystem states, services, and natural capital with quantified uncertainty under global change (J. S. Clark et al., 2001).

In this study, we aim to outline possible ways forward toward developing a framework for continental scale terrestrial ecosystem reanalysis. For this, we review briefly the state-of-the-art of reanalysis in Earth system sciences (Section 2). Section 3 discusses model-data-fusion methods that have been used in Earth system sciences and reanalysis. In Section 3 we also address the computational requirements for long-term data storage and the reanalysis. Section 4 formulates a roadmap toward "terrestrial ecosystem reanalysis" as a short-term objective, outlining a few opportunities. This review also intends to stimulate cross-disciplinary community building involving ecologists, hydrologists, meteorologists, and soil scientists toward a stronger coupling of water, energy and biogeochemical cycles in terrestrial and ecosystem modeling.

## 2. Status of Reanalysis in Earth Sciences

Research on reanalysis in Earth systems is growing rapidly (Figure 1a). Increasing computational capabilities, the growing availability of long-term satellite data with global coverage, advancements in model-data fusion methods such as variational and sequential data assimilation and the increasing awareness of the drastic changes in the Earth system related to anthropogenic and climatic factors drive reanalysis development. This is not only the case for atmospheric sciences, but more recently and more rapidly in hydrologic and ecosystem research, here in the sense of combining reanalysis with soil, hydrologic, biodiversity, and carbon research (Figure 1a). The analysis of recent scientific publications indicates an above-average increase in publications in hydrologic-, carbon cycle-, soil-, and vegetation reanalysis. Reanalysis started in atmospheric science since the launch of the first satellite missions in the 1960s and 70s, succeeded by more than 12,839 publications in meteorology and atmospheric sciences alone. Although atmospheric research still accounts for 43% of published research on reanalysis, research on other Earth system compartments adopted the reanalysis concept to generate continuous, consistent time series used to detect trends, and analyze past states (Figure 1b). Existing reanalysis approaches focus either on specific compartments (atmosphere, land, and ocean) of the Earth system including the interface such as vegetation, or on matter cycles such as carbon, water, and nutrients (Figure 2). Internal feedbacks in the geophysical system are enclosed in a data assimilation framework. Reanalyzes considered in this review are ideally global-scale or continental-scale high-resolution products. They represent a temporally high-resolution description of a coherent geophysical system. Data in Table 1 gives an overview on recent exemplary reanalysis products with a length of at least 5 years and the data being available under FAIR (Findable, Accessible, Interoperable, and Reusable) principles (M. D. Wilkinson et al., 2016). Details of the methodological approach for a data assimilation system are described in Section 3. The reanalysis product depends on the model complexity and data availability (Figure 3). In the next section, we summarize the main achievements and the reanalysis concepts used in the different components of Earth sciences with emphasis on the most recent developments in relation to ecosystem science.

### 2.1. Atmospheric Reanalysis

Atmospheric reanalysis refers to the consistent incorporation of various observations into a numeric weather prediction model using sequential or, more typically, variational data assimilation techniques, with the aim of providing best possible estimates of past atmospheric states (e.g., Kalnay et al., 1996; S. Uppala et al., 2008; Wahl et al., 2017). An important aspect of reanalysis products is that the model system and

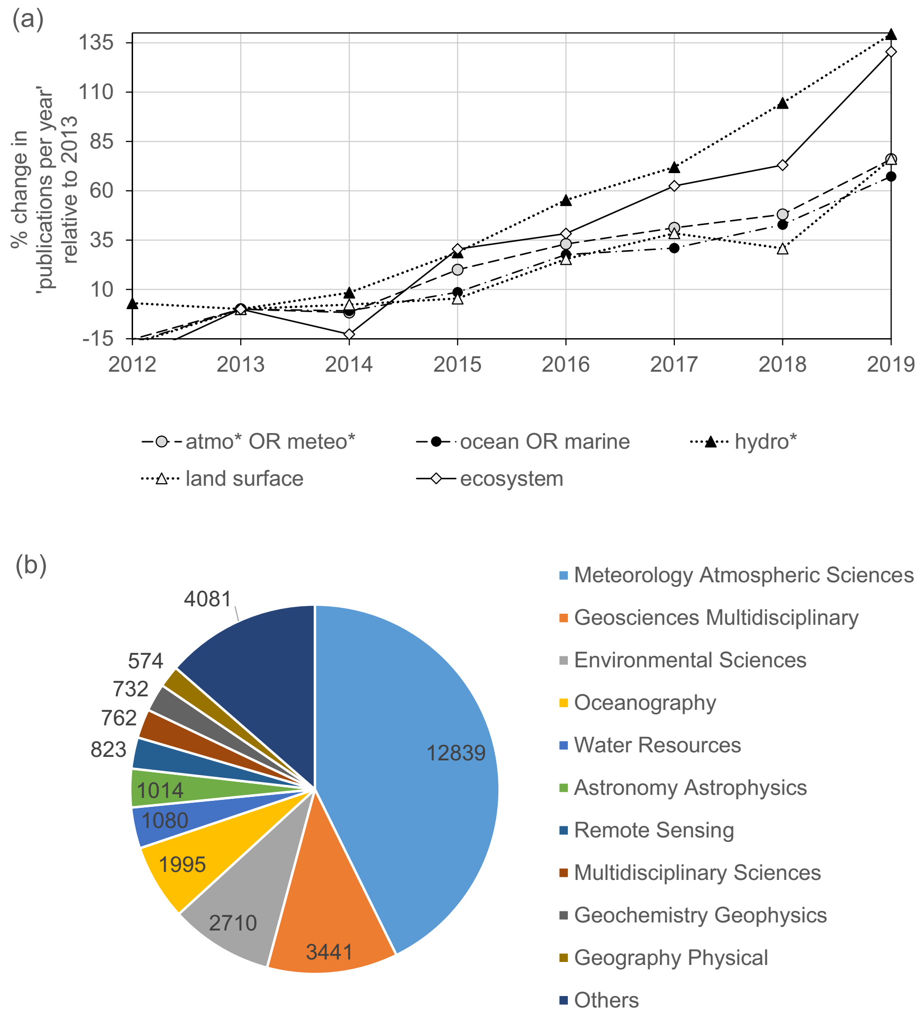


**Figure 1.** Literature review on reanalysis in different Earth system compartments. Results are based on keyword search in https://apps.webofknowledge.com on February 26, 2021. Panel (a) shows publication growth rate of individual science domains from 2013 to 2019, relative to the year 2013. Highest growth rates occurred for hydrological research (139%, 316 publications in 2019, keywords "hydro × and reanalysis"), and ecosystem research (130%, 325 publications in 2019, keywords "(ecosystem OR carbon OR vegetation OR biodiversity OR soil) AND reanalysis"). Rates outperform those of total publication output (30%, not shown) and that of "reanalysis" alone (48%, not shown). Panel (b) shows publications associated to "reanalysis" and the "Web of Science Categories", respectively.

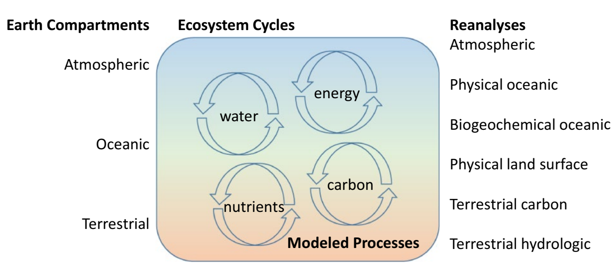


**Figure 2.** Concept of existing reanalysis approaches relative to earth system compartments and physical (water and energy) and biogeochemical ecosystem cycles (water, energy, nutrients, and carbon). Reanalyzes at process level beyond individual compartments is weak, hence remains a scientific challenge.

the data assimilation procedure do not change during the time period of the reanalysis. In contrast, uncertainty characteristics of the assimilated measurement data do change, which is related to the changing technical specifications of observation systems deployed. Currently 7–9 million observations are assimilated at each time step in global atmospheric reanalysis systems and include for example, observations made at the dense global network of meteorological stations, ships, and buoys, information from a large number of satellites, vertical soundings and aircraft.

In the US, the National Center of Environmental Prediction (NCEP) and National Center for Atmospheric Research (NCAR) collaborate on the provision of atmospheric reanalyzes (See Table 1). They produced a first global reanalysis for the period 1957–1996 (Kalnay et al., 1996), which was extended later for the period since 1948 and is still updated, covering now the period 1948–2018. This reanalysis is produced with the global

**Table 1**
*Overview of Exemplary Reanalysis Products in the Earth System Domains and Key Characteristics (See Excel)*

| | Reanalysis product | Lead organization | Data portal | Supporting entity | Time period | Spatial resolution | Spatial extent | Assimilation method | Reference | Vertical Resolution |
|---|---|---|---|---|---|---|---|---|---|---|
| Oceanic | ORAS5 | ECMWF | Yes | University of Hamburg ICDC | 1979–present | 1/4° | Global | 3D-Var | Zuo et al. (2019) | 75 levels |
| Oceanic | SODA3 | NCAR/NOAA/NODC | No | Authors | 1979–present | 1/4° | Global | Optimal interpolation | Carton et al., (2018) | 50 levels |
| Oceanic | ECCOV4.3 | MIT | Yes | NASA Earth Data | 1992–2015 | Variable | Global | 4D-Var | Forget et al. (2015) | 50 levels |
| Oceanic | GLORYS12V1 | Mercator Ocean | Yes | Copernicus | 1993–2019 | 1/12° | Global | SEEK+3DVar | Drevillon et al. (2018) | 50 levels |
| Oceanic | CERA-20C | ECMWF | No | ECMWF | 1900–2010 | ~120 km | Global | Variational | Laloyaux et al. (2018) | variable |
| Oceanic | NWSOC-REA | PML-NCEO | Yes | PML Data Portal | 1998–2009 | ~12 km | Regional | EnKF | Ciavatta et al. (2016) | 42 levels |
| Oceanic | MedPFT-REA | PML-NCEO | Yes | PML Data Portal | 1998–2014 | 1/10° | Regional | EnKF | Ciavatta et al. (2019) | 42 levels |
| Oceanic | NWSHELF_MULTIYEAR_BGC_004_011 | Met Office | Yes | Copernicus | 1993–2019 | ~1/10° | Regional | 3D-Var | Skakala et al. (2018) | 24 levels |
| Oceanic | MEDSEA_REANALYSIS_BIO_006_008 | OGS | Yes | Copernicus | 1999–2018 | 0.06° | Regional | 3D-Var | Teruzzi et al. (2019) | 72 levels |
| Atmospheric | JRA-55 | JMA | Yes | NCAR, DIAS | 1958–2013 | 55 km | Global | 4DVar | Kobayashi (2015) | 60 levels |
| Atmospheric | ERA5 (successor to ERA-Interim) | ECMWF | Yes | CDS Copernicus | 1979–2019 | 30 km | Global | 4D-Var | Hersbach et al. (2020), Dee et al. (2011) | 137 levels |
| Atmospheric | ERA-20C | ECMWF | No | ECMWF | 1899–2009 | 125 km | Global | 4D-Var | Poli et al. 2015 | 37 atm levels |
| Atmospheric | 40 years Reanalysis Project | NCEP/NCAR | No | ECMWF | 1957–1996 | 210 km | Global | 3DVar | Kalnay et al. 1996 | 28 atm levels |
| Atmospheric | CSFR | NCEP/NCAR | Yes | NCAR | 1979–2010 | 0.5° | Global | 3DVar (modified) | Saha et al. 2010 | 64 atm levels |
| Atmospheric | 20CRV3 | NOAA–CIRES–DOE | Yes | NOAA | 1836–2015 | 1° | Global | EnKF | Compo et al. (2011), Slivinski et al. (2021) | 64 levels |
| Atmospheric | Arctic System Reanalysis | Ohio State University | Yes | Authors | 2000–2012 | 15 km | Regional | 3DVAR | Bromwich et al. (2018) | 71 atm levels |
| Atmospheric | COSMO-REA2 | DWD | No | DWD | 2007–2016 | 2km | Regional | Continuous Nudging | Wahl et al. (2017) | 50 atm levels |
| Atmospheric | COSMO-REA6 | DWD | No | DWD | 1995–2017 | 6km | Regional | Continuous Nudging | Bollmeyer et al. (2015) | 40 atm levels |
| Atmospheric/land surface | MERRA-2 (successor to MERRA and MERRA/Land) | NASA | Yes | MDISC–GES | 1979–present | ~0.5° | Global | 3DVAR | Gelaro et al. (2017), Rienecker et al. (2011) | 72 levels |

10.1029/2020RG000715

**Table 1**
*Continued*

| | Reanalysis product | Lead organization | Data portal | Supporting entity | Time period | Spatial resolution | Spatial extent | Assimilation method | Reference | Vertical Resolution |
|---|---|---|---|---|---|---|---|---|---|---|
| Terrestrial (land surface) | GLDAS | NASA/GSFC/ NOAA | Yes | MDISC–GES | 2003–present | 0.25°–2.5° | Global | EnKF | Rodell et al. (2004), Kumar et al. (2019) | 14, plus snow/ veg. |
| Terrestrial (land surface) | ERA5/Land | ECMWF/ Copernicus | Yes | CDS Copernicus | 1981–present | 9km | Global | 4DVar | Muñoz Sabater (2019) | 4, plus snow/ veg. |
| Terrestrial (land surface) | ESSMRA | Forschungszentrum Jülich | Yes | Authors | 2000–2015 | 3km | Regional | EnKF | Naz et al. (2020) | 10 levels |
| Terrestrial (land surface/ carbon) | SMAP Level-4 | NASA/GMAO | Yes | NASA | 2015–present | 9km | Global | Distributed EnKF | Reichle et al. (2016, 2019); Jones et al. (2017) | – |
| Terrestrial (carbon) | CARDAMOM | University of Edinburgh | Yes | Authors | 2001–2010 | 1° | Global | Adaptive Metropolis MCMC | Bloom and Williams (2015) | – |
| Terrestrial (carbon) | CARDAMOM pantropical retrievals 2000–2015 | University of Edinburgh | Yes | Authors | 2000–2015 | 1° | Regional | Adaptive Metropolis | Exbrayat and Willams (2018) | – |
| Terrestrial (hydrology) | Global Water Cycle Reanalysis | Center for Water and Landscape Dynamics | No | Authors | 2003–2012 | 1° | Global | own approach | van Dijk et al (2014) | – |
| Terrestrial (hydrology) | Global Water Cycle | NASA | Yes | GESDISC | 2000–2010 | 1° | Global | own approach | Rodell et al. (2015) | – |
| Terrestrial (hydrology) | GRUN | ETH Zürich | No | Authors | 1902–2014 | 0.5° | Global | MCMC (Data Fusion) | Ghiggi et al. (2019) | – |
| Terrestrial (hydrology) | Continental Runoff into the Oceans | University of Washington | No | Authors | 1950–2008 | 1° | Global | own approach (Data Fusion) | Clark et al. (2015) | – |
| Terrestrial (hydrology) | DOLCE | CCRC | Yes | NCI Australia | 2000–2009 | 0.5° | Global | own approach (Data Fusion) | Hobeichi et al. (2018) | – |
| Terrestrial (hydrology) | CDR | Princeton University | Yes | NCAR | 1984–2010 | 0.5° | Global | own approach (Data Fusion) | Zhang et al. (2017) | - |
| Terrestrial (hydrology) | GRACE-REC | ETH | No | Authors | 1901–2019 | 0.5 | Global | MCMC (Data Fusion) | Humphrey and Gudmundsson (2019) | – |

Abbreviations: CARDAMOM, Carbon Data Model Framework; ECMWF, European Center for Medium-range Weather Forecasts; ESSMRA, European Surface Soil Moisture ReAnalysis; GRACE, Gravity Recovery and Climate Experiment; JMA, Japanese Meteorological Agency; MCMC, National Center for Atmospheric Research; NCAR, National Center for Atmospheric Research; NCEP, National Center of Environmental Prediction; ORAS5, Ocean ReAnalysis System 5; SMAP, Soil Moisture Active Passive;

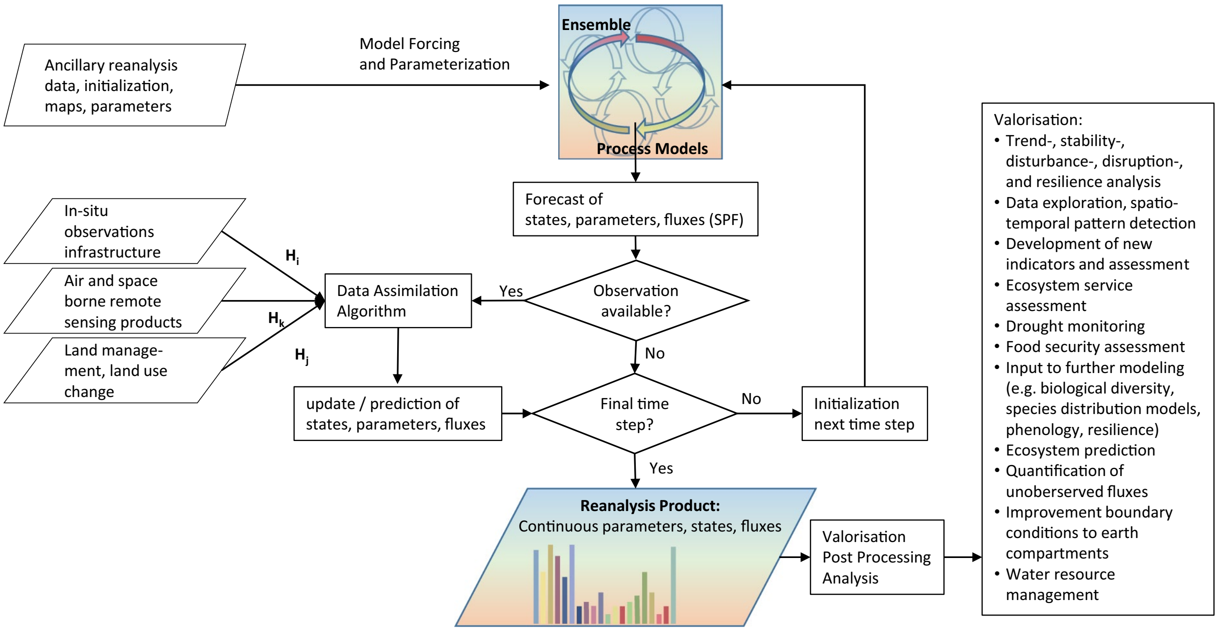


**Figure 3.** Flow chart on creation and valorization of an Earth system reanalysis product, model-data-fusion scheme, uncertainty propagation, and time-step iterative data assimilation.

circulation model Global Forecast System (GFS) and 3D variational data assimilation (3DVAR), at a spatial resolution of 2.5° and for 28 atmospheric levels. Later the so-called Climate Forecast System Reanalysis (CSFR) was published by NCEP, which provides the reanalysis products at a much higher spatial resolution of 0.5° and 64 atmospheric levels for the period 1979–2010 (Saha et al., 2010). The main innovation of this new product is that it considers the coupled atmosphere-ocean-ice-land system, and includes variations in $CO_2$-concentrations, aerosol concentrations and solar activity (Saha et al., 2010). The data assimilation itself is uncoupled which means that measurement-based updates for the atmospheric and ocean compartments of the model are made independently. The modified 3DVAR assimilation system adopted a new model error variances rescaling (Kleist et al., 2009), acknowledged trends in observational data through first-order temporal extrapolation and had quality control algorithm implemented on innovation and variances (Andersson & Jarvinen, 1999).

Regional atmospheric reanalysis can go to even higher resolutions with potential advantages to coarse scale global reanalysis. The Central European COSMO_REA2 operates at 2 km horizontal grid resolution (Wahl et al., 2017), performing convective scale atmospheric prediction by assimilating surface and radar precipitation. ERA-Interim drives the COSMO_REA2 reanalysis as is the case for the COSMO_REA6 reanalysis (Bollmeyer et al., 2015). COSMO_REA6 spans a larger time span while assimilating data from radiosondes, aircrafts, wind profiler, and synoptic surface observations. Both central European reanalyzes show improved representation of high resolution precipitation prediction compared to ERA-Interim. The regional Arctic System Reanalysis (ASR) focused on improving representation of physical processes over frozen surfaces using the Polar Weather Research and Forecasting model (WRF, Powers et al., 2017) with ERA-Interim again providing lateral and initial conditions. Operated at 15 km horizontal resolution, ASR assimilates a multitude of on-ground, air, sea, and satellite observations during the time period 2000–2012. The updated model physics such as subgrid-scale cloud fraction interaction improved correlation and reduced biases in short wave and long wave radiation, seasonal arctic precipitation, and 10 meter wind speed observations (Bromwich et al., 2018).

The European Center for Medium-range Weather Forecasts (ECMWF) published its second global atmospheric ECMWF Re-Analysis (ERA40, S. Uppala et al., 2008) in 2004. The successor of the 15 years ERA15 reanalysis (Gibson et al., 1997) was produced for a 45 year period (1957–2002) at a resolution of 1.125° and for 60 atmospheric levels, using its ECMWF-model in combination with 3DVAR. It was the first atmospheric reanalysis where low product level satellite data were directly assimilated instead of variable retrievals (S.

Uppala et al., 2008). An improved product, ERA-Interim, followed relatively soon thereafter, providing reanalysis at a higher spatial resolution (0.75°), and improved data assimilation with 4DVAR (Dee et al., 2011). Recently, a new ECMWF-reanalysis product, ERA5, has been made available, which is at yet another higher spatial resolution of 0.25° (Hersbach et al., 2020). The main innovation is that coupled data assimilation for the atmosphere-land is used. In addition to atmospheric observations, ERA5 assimilates remotely sensed soil moisture from the ASCAT satellite, snow information, and screen level meteorological measurements. The ERA5 reanalysis is available in near real-time with a few days delay. The near real time data is released as ERA5T data which becomes ERA5 data after 2–3 months lag time. The lag time ensures an option for intervention in the rare event that errors are identified in the near real time window. Some studies showed that the ERA5 reanalysis product outperforms other reanalysis products, for example, on the basis of comparison with precipitation datasets (e.g., Beck et al., 2019). Another important product is ERA-20C, an atmospheric reanalysis for the complete 20th century until 2011, at 125 km (~1.25°) spatial resolution, and also produced with 4DVAR (Poli et al., 2015). In this product no upper atmosphere information and satellite information was assimilated, as this information was not available in the beginning of the 20th century. ERA-20 C does therefore not provide the best estimates of the atmospheric states in the late 20$^{th}$ century and beginning of the 21st century, as for this more recent period ERA5 is a better product. Within this context, NOAA-CIRES-DOE Twentieth Century Reanalysis (20CRv3) project reconstructed four-dimensional global atmospheric states and associated uncertainties using an 80 member ensemble Kalman filter providing three-hourly data from 1806 to 2015 (Compo et al., 2011; Slivinski et al., 2019). Assimilated observations of the era before 1900 are Arctic sea ice extent and sea level pressure. The century long reanalyzes provide means to analyze recent trends in atmospheric circulation patterns in a historical perspective.

Another global reanalysis, the Japanese 55-years Reanalysis JRA-55, is provided by the Japanese Meteorological Agency (JMA), for the period since 1957 and until the near present (with a few days delay). JRA-55 combines the JMA forecast model and 4DVAR (Harada et al., 2016; Kobayashi et al., 2015). JRA-55 succeeded the JRA-25 which provided a reanalysis from 1979 to 2004 (Onogi et al., 2007). Three products are delivered in the context of JRA-55: besides the full-blown reanalysis also a reanalysis is made using only conventional observations and an open loop run without data assimilation (Kobayashi et al., 2015). Finally, MERRA-2 is a global atmospheric reanalysis produced by NASA global modeling and assimilation office (Gelaro et al., 2017). A key strength is the high spatial (0.5°, 72 atmospheric levels) and temporal (hourly) resolution at the global scale and the fact that it is oriented more toward an Earth System Reanalysis, with better representation of land and sea ice processes. For further details on MERRA-2 development see Section 2.3.

Despite the enormous advances in atmospheric reanalysis, the products still have some important limitations. One main problem is the presence of temporal discontinuities due to production stream transitions and changes in the observing systems. This implies, for example, that the detection of trends from reanalyzes must be interpreted cautiously (B. Y. Chen & Liu, 2016; Hobbs et al., 2020; Kossin, 2015). Another main problem is the strong dependency on parameterizations resulting, for example, in unreliable rainfall estimates in convection-dominated regions (Beck et al., 2019). Important ecosystem variables such as soil moisture are considered as tuning variables and outcome of atmospheric processes, which is one reason for often well represented trends but significant differences in magnitude between soil moisture data from reanalysis and remotely sensed or in-situ observations (Rotzer et al., 2015; Yang et al., 2020).

### 2.2. Ocean Reanalysis

Ocean reanalyzes assimilate historical marine observations into models that represent a range of processes from hydrodynamic and sea ice dynamics to biogeochemical reactions and complex trophic interactions of the marine food web. Ocean reanalysis is of particular interest in this review, because ocean reanalysis has progressed further in including ocean ecosystem variables compared to terrestrial reanalysis. Historically, ocean reanalysis has been synonymous with retrospective analysis of physical properties (e.g., water temperature, salinity, current velocities, Penny et al., 2019) using concepts and methods from atmospheric sciences, with the main objective of investigating climate signals and feedbacks (Bengtsson & Shukla, 1988). Since the seminal decadal reanalysis (1982–1992) of sea surface temperature in the Pacific by (Ji et al., 1994), the delivery of ocean physics reanalysis products has become well established in many operational and research centers (Balmaseda et al., 2015). Current state-of-the-art physical ocean reanalyzes include (see Table 1 and Carton et al., 2018) ORAS5 (Ocean ReAnalysis System 5 by the European Center for Medium-Range Weather Forecasts, ECMWF; Zuo et al., 2019), SODA3 (Simple Ocean Data Assimilation, version 3; Carton et al., 2018), ECCO4r3 (Estimating the Circulation and Climate of the Ocean, version 4, release 3; Forget et al., 2015) and GLORYS12V1 (Global Ocean reanalysis and Simulation, version 2 by Mercator Ocean; Lellouche et al., 2018).

ORAS5 is a reconstruction of the ocean and sea-ice state extending back to 1958. ORAS5 was produced with the global ocean-sea ice model with a 1/4° horizontal resolution. The system uses a 3D-Var system to assimilate 5-daily satellite sea surface temperature (SST), sea level anomalies (SLA), sea ice concentration and subsurface temperature and salinity profiles.

SODA3 is a reconstruction of the ocean and sea-ice state that covers the period from 1980 onwards. SODA3 was produced using the global Modular Ocean Model component of the GFDL CM2.5 coupled model and includes an interactive sea ice model. An optimal interpolation system assimilates 10-daily satellite SST and subsurface temperature and salinity profiles.

ECCO4r3 is a reconstruction of the ocean and sea-ice state that covers the period from 1992 to 2015 using the MITgcm. ECCO4r3 employs a 4D-Var method to assimilate SLA, surface salinity, and time-dependent gravity.

GLORYS12V1 reconstructs the ocean and sea-ice state for 1993–2019, characterized by a particularly high spatial resolution (1/12°). It was produced using the NEMOv3.1 ocean model coupled to the LIM2 sea-ice model and used the Singular Extended Evolutive Kalman (SEEK) filter with 3DVar bias correction to assimilate SLA, sea ice concentration, SST, and in situ profiles of temperature and salinity.

Recent advances in coupled data assimilation systems led to reanalysis activities beyond just the physical ocean state. For example, CERA-20C (Laloyaux et al., 2018) is a coupled Earth system reanalysis of the 20th century (1901–2010) reconstructing past weather and climate in the atmosphere, as well as the state of the ocean, land, ocean waves, and sea ice. The model couples the ECMWF's Integrated Forecasting System (IFS) for the atmosphere, land, and waves to the NEMO model for the ocean and to the LIM2 model for sea ice. A variational method with a common 24-h window shared by the atmospheric and ocean components is used to assimilate SLA and marine wind observations as well as ocean temperature and salinity profiles.

Reanalyzes of the ocean chemical and biological components have become available only more recently because of the many unknowns in ecosystem functioning, the sparsity of relevant ocean observations, and high non-linearity of the model equations that challenges traditional Gaussian assumptions in data assimilation methods (Fennel et al., 2019). Early work on biogeochemical data assimilation by (Ishizaka, 1990) capitalized on the availability of large-scale and relatively high-frequency ocean color observations. It was almost 20 years later that the first multi-annual biogeochemical reanalyzes were produced by assimilating ocean-color total chlorophyll in the global ocean (Nerger & Gregg, 2008), in an ocean basin (Fontana et al., 2013) and in coastal and shelf-seas ecosystems (Ciavatta et al., 2016; Hu et al., 2012). More recent contributions include the decadal global ocean ecosystem reanalyzes by Ford and Barciela (2017), obtained by assimilating two different ocean color products for 1997–2012, and the one by Gregg and Rousseaux (2019), who estimated global trends of primary production by assimilating ocean color for 1998–2015.

Besides the well-established assimilation of total chlorophyll from ocean-color (e.g., Hu et al., 2012), innovative applications have assimilated surface ocean color products for: spectral diffuse attenuation coefficients (Ciavatta et al., 2014), size-fractionated chlorophyll and particulate organic carbon (POC; Xiao & Friedrichs, 2014), remote sensing reflectance (E. M. Jones et al., 2016) and both phytoplankton functional type chlorophyll and spectral absorption (Ciavatta et al., 2018, 2019; Pradhan et al., 2019; Skakala et al., 2020). Surface data of partial pressure of $CO_2$ ($pCO_2$) from ships of opportunity were used in the reanalysis of air-sea $CO_2$ fluxes in the global ocean (While et al., 2012). For the ocean interior, biogeochemical simulations were improved by assimilating vertical observations of nutrients, oxygen, and $pCO_2$ data at fixed stations (Allen et al., 2003; Gharamti et al., 2017), glider data of chlorophyll and POC (Kaufman et al., 2018) and biogeochemical-Argo float profiles of oxygen, chlorophyll, photosynthetically available radiation, phytoplankton biomass, and POC (Cossarini et al., 2019; Terzic et al., 2019; Verdy & Mazloff, 2017;

B. Wang et al., 2020). Thus far only few applications addressed weakly versus strongly coupled assimilation of both biogeochemical and physical data but showed these helped to preserve the consistency between physical and biogeochemical structures (Song et al., 2016; Yu et al., 2018). The integrated assimilation of both physical and biogeochemical observations from both satellite and in situ platforms is an active area of research and the likely way forward in marine ecosystem reanalysis (Skákala et al., 2021).

Currently, ocean ecosystem reanalyzes assimilating biogeochemical data are delivered operationally by several centers of the European Copernicus Marine Environment Monitoring Service: in the North West European Shelf and Mediterranean Sea, by assimilating ocean-color with 3DVar methods (Skakala et al., 2018; Teruzzi et al., 2014), and in the Arctic Sea, by assimilating ocean-color with a deterministic Kalman filter. These reanalyzes are finding applications in Blue Growth, management and policy, besides in the delivery of periodical assessments of the state of the ocean (von Schuckmann et al., 2020).

### 2.3. Land Surface Reanalysis

Land surface reanalysis focusses on land-atmosphere states and exchange, that is, mass and energy fluxes between the land surface, vegetation, and the atmosphere. Atmospheric reanalyzes can also provide modeled states and fluxes at the land-atmosphere intersection, which can be tuned to match synoptic observations at meteorological stations, or use land-surface reanalyzes data as lower boundary condition. Land surface states and fluxes are important boundary conditions to hydrologic, ecologic and atmospheric process models. Frequently assimilated variables in land surface reanalysis products are precipitation, soil temperature, soil moisture, snow extent, depth, and snow water equivalent. Most if not all land surface reanalysis are forced with either regional or global atmospheric reanalysis.

One of the earliest systems aiming to assimilate land surface data is the Global Land Data Assimilation System (GLDAS; Rodell et al., 2004). Different GLDAS products exist where the terrestrial processes are modeled with different land surface models: the Noah land surface model (Niu et al., 2011), the Community Land Model version 2 (Bonan et al., 2003), Catchment land surface model (Koster et al., 2000; R. H. Reichle et al., 2011) and the Variable Infiltration Capacity model (VIC; Liang et al., 1994). Although developed for data assimilation, only GLDAS version 2.2 assimilates total water storage of the Gravity Recovery and Climate Experiment satellite mission (GRACE) in the Catchment model (Li et al., 2019). A few of the most common land surface models used in terrestrial reanalysis and data assimilation studies, and represented processes are characterized in Table 2. All land surface models in GLDAS are parameterized by satellite based remotely sensed maps of vegetation, land cover classes, leaf area index, soil properties, elevation and slope. Leaf area index climatology and vegetation parameters are based on 1 km satellite derived information, fractions determine subgrid variability and serve as input to the land surface model run at vertically resolved sub-grid columns.

The Modern-Era Retrospective Analysis for Research and Applications (MERRA; Rienecker et al., 2011) was enhanced toward the land surface reanalysis MERRA-Land by R. H. Reichle et al. (2011) through scaling in-situ precipitation observations with MERRA precipitation, and the revised land surface parameters of the Catchment land surface model (Koster et al., 2000; R. H. Reichle et al., 2011). The Catchment model simulates horizontal surface, subsurface, and runoff water fluxes based on sub-grid topography, bulk soil moisture, land surface heterogeneities, resolved runoff generating processes in addition to the vertical physical water transport which is common to the regular grid based land surface models. The MERRA-Land particularly focused on improving precipitation information for hydrologic land surface and water budget applications. Rienecker et al. (2011) identified the most pressing challenges being the improved representation of surface fluxes and precipitation, and the reanalysis product itself being strongly sensitive to observations. MERRA-Land therefore uses a larger amount of in-situ precipitation to interpolate at grid scale. These findings guided the development of MERRA-2, the successor to MERRA-Land and MERRA. MERRA-2 yields an improved reanalysis of precipitation due to the larger amount of land surface data assimilated, and more abundant hydrologic processes considered in the Catchment land surface model (R. H. Reichle, Draper, et al., 2017).

ECMWF also developed the ERA-Interim/Land reanalysis product (Balsamo et al., 2015). ERA-Interim/Land builds upon the ERA-Interim atmospheric reanalysis by additionally incorporating in-situ precipitation measurements in the assimilation system, and improved land surface parameterization forced by the offline atmospheric reanalysis. Land surface variables such as soil temperature, soil moisture, and snow depth were not assimilated yet but mostly used for assessing the reanalysis skill. The recent ERA5/land reanalysis product from ECMWF, where weakly coupled atmosphere-land data assimilation was performed, assimilates satellite surface soil moisture, and snow cover information (Hersbach et al., 2020). The ERA5 reanalysis is downscaled to 9 km for the ERA5/Land product and incorporates an updated model parameterization for the Hydrology-Tiled ECMWF Scheme for Surface Exchanges over Land (HTESSEL; van den Hurk & Viterbo, 2003) for the period 2000–2019. Similar to previous reanalysis, the ERA5/land uses a multi-year satellite based vegetation climatology (Boussetta et al., 2013), soil mass and energy fluxes, and a snow model (Dutra et al., 2010). A lapse-rate correction maps the output of the ERA5 atmospheric reanalysis to the altitude and interface of the land surface forcing.

Another operational global scale land surface reanalysis that initially was conceived to add value to the Soil Moisture Active Passive (SMAP) mission satellite mission, is the SMAP Level 4 surface and root-zone soil moisture product (R. Reichle et al., 2016; R. H. Reichle et al., 2019). SMAP L4_SM assimilates the SMAP observations into the Catchment land surface model. Being only available since the launch of the SMAP mission in 2015, the SMAP L4_SM product represents the global terrestrial reanalysis with a high spatial resolution. The SMAP L4_SM estimates are further used to estimate the global carbon budget (L. A. Jones et al., 2017) making the SMAP L4 one of only a few linked terrestrial land surface-carbon cycle reanalyzes (see also Section 2.5).

Recently, Naz et al. (2020) provided a 16 years (2000–2015) high resolution (3 km) European Surface Soil Moisture ReAnalysis (ESSMRA). An important aspect is the scaling of coarse observations by the data assimilation system with benefits also for the whole soil moisture profile and runoff simulation (Naz et al., 2020). The assimilation system uses the Ensemble Kalman Filter assimilating coarse resolution (25 km) remotely sensed ESA CCI soil moisture (Gruber et al., 2019; Wagner et al., 2012) into the community land model CLM version 3.5 (Oleson et al., 2008) for generating a 20 member ensemble of soil moisture states. The land surface model used to build the reanalysis is forced with the regional high resolution atmospheric reanalysis COSMO-REA6 (Bollmeyer et al., 2015) and annually variable leaf area index.

Multiple recent studies evaluated and compared land surface and soil moisture reanalysis products, indicating substantial error and therefore recommending great care in using land surface reanalysis products (Rotzer et al., 2015; Ullah et al., 2018). Temporal dynamics are often well represented comparing land surface reanalysis with satellite retrieved soil moisture and in-situ observations. Although temporal dynamics are well represented, absolute values rarely match across reanalysis products (Rotzer et al., 2015). Using soil moisture data from a reanalysis product and ground stations, Deng et al. (2020) reported that changes in soil moisture throughout the karst region of China and its subareas were mainly affected by precipitation, followed by temperature. In summary, climate, vegetation, and geological background determined the spatiotemporal distribution of soil moisture. The study found that the soil drying trend in recent decades and global climate change are not conducive to the ecological restoration of vulnerable karst areas. Y. W. Wang et al. (2021) conducted a global assessment of two blended microwave soil moisture products using in-situ measurements from the International Soil Moisture Network. The results indicated that the product from the Climate Change Initiative of the European Space Agency reveals overall better accuracy than that of the Soil Moisture Operational Product System from the National Oceanic and Atmospheric Administration. Ullah et al. (2018) reported significant errors in the soil moisture of reanalysis products particularly for freezing-thawing conditions given identical data input and different land surface models. Frequently, land surface reanalysis products are also compared in hydrologic studies (e.g., Ndehedehe et al., 2018) and to in-situ and remote sensing observations (e.g., Albergel et al., 2013; Kumar et al., 2019; Su et al., 2018). We note that the update of biogeochemical soil and vegetation states, plant phenology and parameters were not considered in these land surface reanalyzes. Carbon cycle reanalysis allows for resolution of biogeochemical processes beyond climatology parameterization. Terrestrial hydrologic reanalysis enhances the resolution of the terrestrial water cycle using enhanced horizontal connectivity and additional data sources.

**Table 2**
*Models Potentially Suitable for Terrestrial Ecosystem Reanalysis, Key Processes and Parameterization Modes*

| Land surface and ecosystem model | Land cover/ vegetation Classes | Surface representation | Soil temperature | Canopy hydrology | Soil hydrology |
|---|---|---|---|---|---|
| CARDAMOM-DALEC | | One canopy layer, and three soil layers | N/A | interception, throughfall and drip | Surface runoff, infiltration, sub-surface drainage, redistribution of water within soil |
| CHTESSEL | 20 | Two vegetation layer, 4 soil layers, 1 snow layer | Heat transfer in soil and snow | interception, throughfall and drip | Surface runoff, infiltration, sub-surface drainage, redistribution of water within soil |
| CLM (3.5) | 17 | One vegetation layer, 10 soil layers, 5 snow layers | Heat transfer in soil and snow | interception, throughfall and drip | Surface runoff, infiltration, sub-surface drainage, redistribution of water within soil, 1-D groundwater |
| CLM(5.0) | 16 | One vegetation layer, 25 soil layers, 10 snow layers | Heat transfer in soil and snow | interception, throughfall and drip | Surface runoff, infiltration, sub-surface drainage, redistribution of water within soil, 1-D groundwater |
| CLM-PF | 17 | One vegetation layer, 10 soil layers, 5 snow layers | Heat transfer in soil and snow | interception, throughfall and drip | Surface runoff, infiltration, sub-surface drainage, redistribution of water within soil, 3-D groundwater |
| HTESSEL | 20 | Two vegetation layer, 4 soil layers, 1 snow layer | Heat transfer in soil and snow | interception, throughfall and drip | Surface runoff, infiltration, sub-surface drainage, redistribution of water within soil |
| ISBA | | One vegetation layer, 2–14 soil layers, 12 snow layers | Heat transfer in soil and snow | interception, throughfall and drip | Forece-store |
| JULES | 9 | Multiple canopy layers, 4 soil layers, multiple snow layers | | interception, throughfall and drip | Surface runoff, infiltration, sub-surface drainage, redistribution of water within soil, lower boundary no-flux condition |
| LPJ-Guess | 10 | Two vegetation layer, 2 soil layers, no snow layer | No heat transfer | No | Surface runoff, infiltration, sub-surface drainage, redistribution of water within soil |
| mHM | 3-16[a] | 6 soil layers; 1 snow layer | Energy balance approximation; | interception, throughfall and drip | surface runoff, infiltration, subsurface runoff contributions (saturation excess), groundwater reservoir[b] |
| Noah-MP | 17 | One canopy layer, three snow layers, and four soil layers | Heat transfer in soil and snow | canopy interception loss, and transpiration, uniformly distributed roots and varying root depths | Surface runoff, infiltration, sub-surface drainage, redistribution of water within soil, 1-D groundwater |
| VIC | 12 | Two vegetation layer, 3 soil layers, 2 snow layers | Heat transfer in soil and snow | interception, throughfall and drip | Surface runoff, infiltration, sub-surface drainage, redistribution of water within soil |

Abbreviations: CARDAMOM, Carbon Data Model Framework; HTESSEL, Hydrology-Tiled ECMWF Scheme for Surface Exchanges over Land; ISBA, Interactions Soil Biosphere Atmosphere; VIC, Variable Infiltration Capacity model.
[a]Depends on the data sets available. [b]Detail groundwater coupled with OpenGeoSys. [c]Land cover dynamically affect model parameters via MPR upscaling (Samaniego et al., 2010).
[d]Via MPR for all input geophysical data (soil properties, DEM, etc) and all parameters. Some processes are also represented at multiple scales.

| Runoff routing | Carbon cycle | Land cover change | Snow hydrology | Subgrid variability | Scale | Grid structure | Reference |
|---|---|---|---|---|---|---|---|
| no | Yes (4 plant pools, 2 dead organic matter pools) | From satellite data | No | Variable resolution grid | global/ regional | regular grid | Bloom et al. 2016 |
| no | Yes | Static | Snow accumulation and melt, compaction and water transfer between snow layers | | global/ regional | regular grid | Boussetta et al. 2013, JGR |
| yes | No | Static | Snow accumulation and melt, compaction and water transfer between snow layers | Subgrid PTFs | global/ regional | regular/irregular grid | Oleson et al. 2007 |
| yes | Photosynthesis, litter C, SOC, decomposition, plant respiration, and disturbance (fire) | Dynamic | Snow accumulation and melt, compaction and water transfer between snow layers | Subgrid PTFs, runoff | global/ regional | regular/irregular grid | Lawrence et al. 2019 |
| yes | No | Static | Snow accumulation and melt, compaction and water transfer between snow layers | Subgrid PTFs | regional | regular/irregular grid | Maxwell and Kollet 2008 |
| No | no | Static | Snow accumulation and melt, compaction and water transfer between snow layers | | global/ regional | regular grid | Balsamo et al. 2009, JHM |
| No | yes | Static | Snow accumulation and melt, compaction and water transfer between snow layers | Subgrid runoff | global/ regional | regular grid | Noilhan and Mahfouf 1996, GPC |
| yes | yes | Yes | zero-layer model and multi-layer snow | Subgrid PTFs | global/ regional | regular grid | Best et al. 2011 |
| No | Photosynthesis, litter C, SOC, decomposition, plant respiration, and disturbance (fire) | Dynamic | No | | global/ regional | regular grid | Smith et al. (2001) |
| multiscale river routing + lake modeling | no | dynamic[c] | snow accumulation and melting | yes[d] | global/ regional | regular/irregular grid | Samaniego et al. 2010 |
| No | yes (plant: 3 pools: leaf, wood, and root) and two soil carbon pools (fast and slow). | Static | liquid water storage in snow, snow-interception model, melt/refreeze capability, sublimation of canopy-intercepted snow | semi-tile vegetation and bare soil | global/ regional | regular grid | Niu et al. 2011 |
| yes | No | Static | Snow accumulation and melt, compaction and water transfer between snow layers | | global/ regional | regular grid | Liang et al. 1994 |

### 2.4. Hydrological Reanalysis

A large number of studies have attempted to quantify aspects of the Earth's hydrological cycle through the combination of multiple data sources by using either data assimilation (which can be considered reanalyzes) or statistical model-data fusion approaches which cannot be considered reanalyzes in the strict sense but we believe are also of interest to the readership. Here, we provide a brief overview of key (quasi-)global studies focusing on the overall water budget (Pan et al., 2012; Rodell et al., 2015; Sahoo et al., 2011; Zhang et al., 2018) or on specific hydrological components, including evapotranspiration (Hobeichi et al., 2018; Jimenez et al., 2018), runoff (E. A. Clark et al., 2015; Ghiggi et al., 2019; Hobeichi et al., 2019), and terrestrial water storage (Humphrey & Gudmundsson, 2019; van Dijk et al., 2014).

Among the four studies focusing on the overall water budget, two used data assimilation to blend multiple data sources (Pan et al., 2012; Zhang et al., 2018), Sahoo et al. (2011) used weighted averaging with the Constrained Ensemble Kalman Filter (CenKF) to close the water balance component, while Rodell et al. (2015) used simple (unweighted) water balance averaging. The two earliest studies (Pan et al., 2012; Sahoo et al., 2011) performed their analyses using catchment averages, Rodell et al. (2015) used continental averages, whereas Zhang et al. (2018) used 0.5°-resolution gridded averages. Sahoo et al. (2011) relied primarily on satellite-based sources for quantifying the four main components of the water budget (precipitation, evapotranspiration, runoff, and terrestrial water storage) for 10 major river catchments globally for a four-year period (2003–2006). They found that closure was generally not possible, with errors ranging from 5% to 25% of mean annual precipitation. Pan et al. (2012) considered the same four components, but analyzed a longer period (1984–2006) and also considered in-situ- and model-based data sources. They focused on 32 major catchments across the globe, and concluded that the assumptions underpinning the error analysis may not hold for all regions. Rodell et al. (2015) focused on the entire globe and analyzed the atmospheric water budget in addition to the surface water budget for the period 2000–2010. They found that closure within 10% was generally possible. Zhang et al. (2018) analyzed the entire land surface for the period 1984–2010 and reported that the main challenge was the sparseness and limited availability of "ground truth" observations for bias correction of the individual components.

The two studies focusing on evapotranspiration (Hobeichi et al., 2018; Jimenez et al., 2018) used FLUXNET observations (http://fluxnet.fluxdata.org/) as reference to derive merged gridded estimates for the land surface at 0.5° resolution for 2000–2009 and at 0.25° resolution for 2002–2007, respectively. Arguably the main challenge in these studies is the fact that evapotranspiration is a largely invisible process that is, highly variable in space and time and therefore difficult to measure. The three studies focusing on runoff (E. A. Clark et al., 2015; Ghiggi et al., 2019; Hobeichi et al., 2019) all used streamflow observations from the Global Runoff Data Center (GRDC) as reference to derive 0.5°-resolution merged gridded estimates using bias-correction for 1950–2008, optimal weighting for 1980–2012, and random forest regression for 1902–2014, respectively. Major obstacles to derive merged gridded runoff estimates are the scale discrepancy between local runoff from individual hillslopes and streamflow from large catchments and the fact that the majority of the globe is ungauged or poorly gauged. The two studies focusing on terrestrial water storage (Humphrey & Gudmundsson, 2019; van Dijk et al., 2014) both used gravity anomaly measurements from the GRACE satellite-pair to obtain gridded terrestrial water storage estimates for 2003–2012 using data assimilation and for 1901–2014 using a statistical approach, respectively. A limitation of GRACE is the lack of fine spatial detail due to GRACE's large footprint size (∼400 km).

Hydrologic connectivity in lateral and vertical direction are key processes in these reanalyzes. Vertical oriented grid based land surface models used for land surface reanalysis often limit lateral flow, ponding and infiltration process representation (Vereecken et al., 2019). Distributed hydrological models integrate surface-subsurface water interaction across nodes and grid cells by solving large systems of partial differential equations including lateral water fluxes, infiltration, unsaturated, surface, and groundwater flow. The Interactions Soil Biosphere Atmosphere (ISBA-SURFEX) model (Decharme & Douville, 2006; Noilhan & Mahfouf, 1996), ParFlow-CLM (Kollet & Maxwell, 2008), and the Catchment Land Surface Model (Catchment LSM) are examples of such distributed or semi-distributed hydrologic models. However, a resolution of above 25 km grid size, model setup (e.g., climatology) and limited data availability undermine the assumptions for three-dimensional water flow process representation in support of one-dimensional vertical water flow process representation with limited grid cell inter-action. Spatial resolution is also a key challenge for dynamic vegetation, nutrient and carbon cycle modeling in terrestrial carbon cycle reanalysis.

### 2.5. Terrestrial Carbon Cycle Reanalysis

This section discusses terrestrial carbon cycle models and their incorporation into carbon cycle reanalysis. Terrestrial carbon cycle models include a detailed representation of land surface biogeochemistry, targeted at carbon cycling in plants, and soils (e.g., photosynthesis and respiration), with a model-dependent element of feedback between the C cycle and climate, water and N cycles. Models vary in their complexity, and can include carbon allocation, leaf phenology, impacts of biomass burning and wildfires, disturbance and succession, land use and land cover change. Vegetation in these models is commonly represented and parameterized by plant functional types. These are groupings of species with similar life forms (e.g., grasses, trees) that show a similar response to a given set of environmental conditions (e.g., tropical, boreal). Terrestrial carbon models rely on meteorological reanalysis data like precipitation, air temperature, shortwave radiation and vapor pressure deficit as model driving inputs for regional to global applications.

The key processes in carbon cycle models have been studied at various scales by networks of eddy covariance measurements (i.e., FLUXNET, Papale et al., 2006; Reichstein et al., 2007), field experiments (Chabbi et al., 2017; Grosse et al., 2020) and ecosystem manipulations simulating for example, atmospheric carbon dioxide enrichment (Ainsworth & Long, 2005; Wieder et al., 2019). However, the terrestrial carbon cycle remains poorly constrained, with uncertain interactions and feedbacks (e.g., Friedlingstein et al., 2014; Huntzinger et al., 2017). This is largely because terrestrial carbon models rely on these *in situ* observational data for model parameterization and verification. Data time series are relatively short, and data are especially scarce from the tropical, arctic, and boreal regions compared to other regions. However, advances in satellite observation platforms, combined with new and existing *in situ* data, provide an opportunity to combine these data with terrestrial carbon models using, for example, data assimilation, for a better understanding of the carbon cycle in these regions and elsewhere (Schimel et al., 2015; Scholze et al., 2017).

Pioneering work was done in the 1990s by Knorr and Heimann (1995) using the Simple Diagnostics Biosphere Model coupled at a 0.5° resolution with a 7.83° × 10.0° atmospheric tracer circulation model. The atmosphere-biosphere interaction was constrained using FAPAR remote sensing and sparse in-situ $CO_2$ observations at five sites over the period of 2 years. Since then reanalyzes of the terrestrial carbon cycle have been produced with a broad range of approaches and increased their resolution and data inputs. The objective of these reanalyzes has been to inform process-based carbon modeling using measurements for calibration, to provide a better estimate of the terrestrial carbon cycle. While the history, methodology, and models used in carbon cycle data assimilation systems are extensively discussed in Scholze et al. (2017), we include key highlights below:

The land component of the Carbon Cycle Data Assimilation System project (CCDAS) was initially built around the Jena Scheme for Biosphere Atmosphere Coupling in Hamburg (Raddatz et al., 2007; Scholze et al., 2017), followed by Biosphere Energy-Transfer Hydrology (BETHY, Knorr, 2000) model with the objective to provide a reanalysis for carbon fluxes (Kaminski et al., 2012). Within CCDAS, FAPAR data, eddy flux data (Baldocchi et al., 2001), and remotely sensed soil moisture data are assimilated, leading to reduced land surface flux uncertainty. The CCDAS assimilation approach uses an adjoint method, whereby a tangent linear version of the model code is generated through automatic differentiation. From this adjoint, parameters can be optimized to the data input.

The CCDAS framework was more recently used with the ORCHIDEE land surface model to generate a carbon reanalysis set for the years 2000–2009 (Peylin et al., 2016). This product assimilated FLUXNET observations, atmospheric $CO_2$ emissions and the Normalized Difference Vegetation Index (NDVI) monitored from the Moderate Resolution Imaging Spectrometer (MODIS) as a proxy for plant phenology, which significantly constrained the growing season. The CCDAS framework has also been applied to assimilate new satellite-derived measurements of solar induced fluorescence to constrain monthly and daily simulated carbon fluxes and parameter values (Koffi et al., 2015), replacing the BETHY model photosynthesis schemes with those from SCOPE (Soil Canopy Observation, Photochemistry, and Energy fluxes), which also includes a fluorescence model (van der Tol et al., 2009).

A recent CCDAS study of Castro-Morales et al. (2019) simulated three decades of global terrestrial carbon fluxes by assimilating data with different periods of observations to evaluate the ability of a carbon cycle reanalysis to predict long-term trends and variability in the global carbon cycle. Here, the data assimilation framework was built upon the JSBACH land surface model (Dalmonech & Zaehle, 2013; Raddatz et al., 2007; Reick et al., 2013) and is called the Max Planck Institute MPI-CCDAS. This framework uses the methodology of Kaminski et al. (2012) that simultaneously reduces the model–data mismatch for multiple independent carbon cycle data sets. Castro-Morales et al. (2019) find MPI-CCDAS is capable of simultaneously integrating two independent observational data sets, the fraction of absorbed photosynthetic active radiation (FAPAR) and atmospheric CO2 concentrations, over three consecutive decades at the global scale to estimate global terrestrial carbon fluxes. MPI-CCDAS can confidently predict carbon fluxes up to five years, with reduced certainty for long-term forecasts.

The Carbon Data Model Framework (CARDAMOM, Bloom & Williams, 2015), uses a different approach to data assimilation, that avoids the need for an adjoint. CARDAMOM uses Monte Carlo methods, where large ensembles of the selected model are simulated and evaluated against observations. Bayesian methods are used to accept or reject parameters such that a set of parameters consistent with data and data errors are identified. Because millions of model runs are required to find robust solutions, the model must be fast running, and therefore of intermediate complexity, for example, DALEC (M. Williams et al., 2005). CARDAMOM also makes use of ecological and dynamic constraints to simplify the search for realistic parameters (Bloom & Williams, 2015). CARDAMOM has been used globally to assimilate biomass and soil maps and time series of LAI data to assess plant carbon allocation, stocks, residence time, and carbon use efficiency. Because CARDAMOM avoids the specification of plant functional types, parameter maps are emergent from the DA process pixel by pixel across the model domain. As a result, a key finding from CARDAMOM is that land cover types used by typical Earth System Models do not adequately depict the spatial variability of carbon cycle parameters and processes (Bloom et al., 2016). CARDAMOM has been used to estimate the global terrestrial carbon cycle for the first 10 years of the 21st Century, identifying biome and continental variations in carbon residence times (Bloom et al., 2016) that are critical determinants of C-climate sensitivity in model forecasts.

Although with currently a relatively short time span starting in March 2015, the SMAP Level-4 Carbon product is an interesting candidate for a carbon-related reanalysis. It is driven by surface meteorological forcing data from the Goddard Earth Observing System (GEOS), the SMAP L4_SM (R. Reichle et al., 2016; R. H. Reichle et al., 2019), and Moderate Resolution Imaging Spectroradiometer (MODIS) satellite observations (land cover classification and 8-days canopy FPAR). From the CASA and CENTURY models the simplified light use-efficiency plant production and soil organic matter decomposition logic have been implemented, relating the carbon cycle components to basic meteorological conditions and vegetation-ecosystem functional characteristics (L. A. Jones et al., 2017). The global results are provided at 9 km and daily resolutions, and inform about Net Ecosystem CO2 exchange (NEE) computed as total respiration (vegetation plus soil) less vegetation gross primary production (GPP). By improving the coupling between the water and carbon cycles, the SMAP Level-4 Carbon product may pave the way toward an integrative hydrology-carbon reanalysis.

Overall, these studies are promising in terms of leading to a better ability to understand the carbon sink strength of terrestrial ecosystems across the globe, while also emphasizing the importance of operational networks for *in situ* data and continuous effort. These studies improve the incorporation of the terrestrial carbon cycle dynamics to consider and quantify its feedbacks to the climate system (Heimann & Reichstein, 2008). There is a strong potential for future work to advance C cycle understanding, given the number of new satellite observations being prepared or planned. C cycle DA should aim to assimilate atmospheric $CO_2$ data (Schimel et al., 2015); direct observations of the C cycle from repeat biomass mapping (Quegan et al., 2019); high resolution and frequent sensing of canopy properties; photosynthesis inferred from solar induced fluorescence (Damm et al., 2010); burned area and land use change mapping (Hansen et al., 2010); soil and canopy moisture (van der Schalie et al., 2016) and evapotranspiration (Hulley et al., 2017).

### 2.6. Error Assessment for Reanalysis

A key challenge for any reanalysis is to have access to robust estimates of error associated with the satellite or other products that are assimilated. The error plays a key role in weighting the inversion process. If the weighting is incorrect then the reanalysis may be biased by products that are of lower quality than expected. To mitigate this risk requires close interaction with field scientists who can provide calibration data for the algorithms used to convert remote sensing or other data into ecological information. This calibration process can attach a clear estimate of systematic and random bias to the satellite product (e.g., see Ryan et al., 2012 for a biomass error characterization). While more satellite products now come with error statistics, independent evaluation indicates that these errors can be underestimated (Y. Zhao et al., 2020). A concerted effect to characterize error robustly is required for effect reanalysis.

## 3. Status of Model Data Fusion Methods in Reanalysis

Bayesian statistics, stochastic algorithms and ensemble simulation methods are frequently used to quantify the uncertainties in model-data-driven reanalyzes ranging from ocean sciences (Axell & Liu, 2016; Ciavatta et al., 2016; Schartau et al., 2010; Simon et al., 2015; J. P. Xie et al., 2017; Xue et al., 2012), atmospheric sciences (Candiani et al., 2013; Compo et al., 2011; Whitaker et al., 2004), to terrestrial ecosystem sciences (Dunne & Entekhabi, 2005; Montzka et al., 2012). The uncertainty of a reanalysis can be quantified and the magnitude of the uncertainty typically depends on observation uncertainty and influence, uncertainty of model inputs (e.g., atmospheric forcings), (uncertain) model dynamics, process representation, soil and vegetation parameter uncertainty, and parameters (Dietze, 2017; J. P. Xie et al., 2017). Reanalysis across various disciplines as described in Sections 2.1–2.5 is based on a variety of model-data fusion methods including variational data assimilation (3DVAR, 4DVAR, and hybrid 4DVAR), sequential data assimilation (various variants of the Kalman filter and the particle filter), smoothing and Markov Chain Monte Carlo methods (MCMC). These methods are discussed in more detail and with limited examples in the following sections. This section focusses more on the methodological aspects while examples for reanalyzes are provided in the reanalysis product Table 1.

### 3.1. Ensemble Kalman Filter

The Kalman filter was originally developed for linear systems (Kalman, 1960). The key of the Kalman filter is the dynamic propagation of the prior estimate errors by a prognostic model. Because ecosystem models and remote sensing observation models are non-linear, and because the prognostic state vector can be large, the ensemble Kalman filter (EnKF) was proposed (Burgers et al., 1998; Evensen, 2003; R. H. Reichle et al., 2002) to diagnose the state error covariance matrix recursively from a sample ensemble of realizations. The Ensemble Kalman Filter is frequently used to generate atmospheric (Compo et al., 2011; Slivinski et al., 2019), oceanic (Ciavatta et al., 2019), and terrestrial reanalyzes (Naz et al., 2020; R. H. Reichle et al., 2019). An ensemble of model forecasts is generated by perturbing model input, state variables and/or parameters. These forecasts are compared to observations, when available. Based on the relative ensemble-based uncertainty in the forecasts and the observations, the ensemble model trajectories are updated and used as initial conditions to subsequent forecasts. The EnKF is directly applicable to estimate unobserved variables, for example, to estimate root-zone soil moisture from observed brightness temperature, to inter- and extrapolate information from one region to another (R. H. Reichle et al., 2014), or to downscale data via spatial filtering. The EnKF is also used for the joint estimation of model states and unknown parameters (Y. Chen & Zhang, 2006). The EnKF is one of the most popular techniques for sequential data assimilation in non-linear systems for its ease of implementation, but it also has shortcomings. In case of a larger state vector, the ensemble size needs to increase; the choice of the perturbation parameters will determine the optimality of the filter; covariance inflation may be needed to keep sufficient ensemble spread, and state variables are assumed to follow a Gaussian distribution, among others. Important assumptions to the EnKF and other filters are prior knowledge about observation and process error covariance matrices. Recent developments include efficient EnKF-variants which compress state space, and improving the performance of EnKF by extending the data assimilation window, and the use of an iterative approach, which gives rise to iterative ensemble smoothers (Bocquet & Sakov, 2014; Emerick & Reynolds, 2013; Evensen, 2018). Also the generalization of the EnKF method toward new assimilation algorithms is a recent development which enables ecosystem models and assimilation of non-Gaussian data (Raiho et al., 2020). Many studies have focused on the sequential assimilation of soil moisture data (M. L. Carrera et al., 2015; De Lannoy & Reichle, 2016; de Rosnay et al., 2013; Draper et al., 2012; R. H. Reichle, De Lannoy, et al., 2017) and snow (De Lannoy et al., 2010; Kumar et al., 2017; Thirel et al., 2013; Toure et al., 2018). M. Williams et al. (2005) used EnKF to integrate both flux data and biometrics (stock data) in a model analysis of a forest carbon cycle. Smoothing has also been used for soil moisture (Dunne & Entekhabi, 2005), to reconstruct snow (Durand et al., 2008) or to infer snow and soil and deep groundwater estimates from GRACE measurements (Girotto et al., 2016; Kumar et al., 2016; Zaitchik et al., 2008).

### 3.2. Particle Filter

A particle filter is essentially a sequential Monte Carlo method that approximates the probability density function of the posterior states by a set of random samples. Compared to the EnKF, the particle filter has the advantage that it can correctly process uncertainties that follow any non-Gaussian distribution. In particle filtering, the probability density function of the posterior state vector is approximated by a set of random samples. These are labeled particles, each of which has its own weight. A first step in the application of the particle filter is Sequential Importance Sampling (SIS), in which each of the weights are updated using the external observations. Due to the use of a sub-optimal approximation of Bayes' rule, the variance of the weights tends to increase, leading to a large number of particles with small importance weights. For this reason, Sequential Importance Resampling (SIR) needs to be applied. In this step, particles with a high importance are retained, and their values are assigned to the state vectors of particles with low weights. The weights of the new set of particles are then all set equal. One problem that may arise is particle degeneration, in which the particle set collapses to a single particle. This can be solved through the resample-move step, in which Monte Carlo Markov Chains are applied to the particle set. For a more detailed description of the Particle Filter we refer to Doucet et al. (2000). Recent developments in particle filtering are more promising for their application in high dimensional systems, and therefore also for a possible future role in reanalyzes. Examples are the variational mapping particle filter (Pulido & van Leeuwen, 2019) and localized particle filter (Penny & Miyoshi, 2016; Poterjoy, 2016). Studies focusing on the application of the particle filter in ecological sciences can be found in (Dowd & Joy, 2011; Knape & de Valpine, 2012; Martin-Fernandez et al., 2014; Rakhimberdiev et al., 2015; Weir et al., 2013). For ecosystem reanalysis we can expect that many states and parameters will show non-Gaussian distributions, so that Kalman-type assimilation algorithms perform suboptimal. Efficient particle filters are therefore an interesting future alternative for ecosystem reanalysis, but currently not efficient enough and not well established for reanalyzes.

### 3.3. Variational Data Assimilation: 3DVAR and 4DVAR

Variational data assimilation is based on a Gaussian approximation of Bayes law, weighting simulation model-data mismatches on one hand (the likelihood term) and mismatches between updated model variables and prior model variables on the other hand (referred to as the background part in atmospheric sciences). This results typically in a two part objective function which can be minimized making use of adjoint state techniques in combination with non-linear optimization (Compo et al., 2011). In Earth sciences, either 3D variational data assimilation (3DVAR) or 4D variational data assimilation (4DVAR) is used. In 4DVAR, the minimization is carried out over a longer time window taking into account that measurements which are used for assimilation are made at different time points. 4DVAR is commonly used in combination with atmospheric models (e.g., see Kalnay, 2003 for an overview), and also in groundwater hydrology (e.g., J. Carrera & Neuman, 1986; Gomez-Hernandez et al., 1997). For land surface data assimilation with a focus on the carbon cycle, 4DVAR is used in combination with the ORCHIDEE-model (Peylin et al., 2016). New methodological developments include the estimation of the background error covariances through calculating the spread of the ensemble in hybrid 4DVAR methods. In hybrid 4DVAR the background error covariance matrix is combined from a prior, static covariance matrix and a covariance matrix from an ensemble. Ensemble 4DVAR methods are based on an ensemble of trajectories so that the use of linear and tangent models at each iteration is avoided, and the model error covariance matrix is calculated from the ensemble (Lorenc et al., 2015). Variational data assimilation is also of high interest for ecosystem reanalysis, as we are dealing in ecosystem reanalysis with processes which act on very different time scales, and many unknown parameters which are also linked to these disparate time scales. Minimization of an objective function over a larger time window can potentially take better account of these different time scales than sequential data assimilation.

### 3.4. Markov Chain Monte Carlo (MCMC)

Markov-Chain Monte-Carlo is a batch method for state and parameter estimation. MCMC solves general inverse problems through determining the probability distribution of a vector of model parameters given a set of measurements (Knorr & Kattge, 2005). The initial approach was developed by Metropolis et al. (1953) to directly sample the parameter probability distribution using Monte Carlo techniques. The complete method of Monte Carlo inversion is described in detail by Mosegaard and Tarantola (1995) and reviewed by Mosegaard and Sambridge (2002). The Metropolis algorithm operates by updating prior information on parameters (expected values or ranges) with a model-observation comparison, to guide a Markov chain (random walk) through parameter space and has been adopted by Vrugt et al. (2013). The algorithm is a batch process, so for each proposed parameter step a complete run of the model is required. Multiple chains, each typically with millions of members, are required to test for consistency in final likelihoods, using convergence tests. The REFLEX experiment (A. Fox et al., 2009) explored the efficacy of varied different data assimilation techniques, including Kalman filters and batch methods like MCMC to analyze carbon flux data. Bloom et al. (2016) used the adaptive Metropolis MCMC approach in combination with a simplified ecological model to produce a reanalysis of the terrestrial carbon cycle. Overall MCMC can have high computational costs due to the large number of simulations required, but unlike the EnKF, MCMC preserves any mass balance imposed by the model (Hill et al., 2012). The MCMC method is not yet efficient enough for large scale ecosystem reanalyzes, but its use in an ecosystem reanalysis could be the estimation of vegetation trait specific parameters at highly equipped sites, which are used at other sites (grid cells) with the same vegetation type.

### 3.5. Software Tools Available for Model Data Fusion

State data assimilation can start simple–the classic Kalman Filter can be implemented in a few lines of code for the scalar case -, but implementation can quickly become more complicated when using more advanced methods that relax simplifying assumptions or deal with large spatial extents or data volumes. On top of this is the informatics hurdle of processing model inputs, starting and stopping ensembles of model runs, and ingesting data constraints, which can quickly spiral out of control. By contrast, well-designed software frameworks can generalize and abstract many of these steps, allowing data assimilation systems to scale more easily. Utilizing community-supported tools can also be more robust, as a larger development community has played a role in testing code, and more efficient, as it results in less duplicated efforts and shallower learning curves (Fer et al., 2020).

Fortunately, when it comes to the state data assimilation approaches used in reanalysis and iterative forecasting, there are a number of alternative frameworks available. One of the most established is the Data Assimilation Research Testbed (DART, https://www.image.ucar.edu/DAReS/DART/) maintained by the U.S.'s National Center for Atmospheric Research (NCAR). DART supports a number of different ensemble-based DA algorithms (Anderson et al., 2009), with a focus on the Ensemble Adjustment Kalman Filter (EAKF), a variant of EnKF that nudges ensemble members. In terms of ecosystem applications, A. M. Fox et al. (2018) coupled DART to NCAR's Community Land Model (CLM4.5) and demonstrated its ability to assimilate both real and simulated biomass and leaf area data at the site-level using observations from an Ameriflux site in New Mexico. In this application ensemble spread was generated based on state and driver uncertainty, and parameters were held constant. Viskari et al. (2015) also coupled DART to the Ecosystem Demography model (ED2, Medvigy et al., 2009) to assimilate both tower-based and remotely sensed (MODIS) phenological observations. Similar to DART, the German Alfred-Wegener-Institute maintains the Parallel Data Assimilation Framework (PDAF, http://pdaf.awi.de/), which similarly supports multiple ensemble-based algorithms (Nerger & Hiller, 2013). Kurtz et al. (2016) coupled PDAF to the terrestrial systems modeling platform (P. Shrestha et al., 2014) achieving a very favorable scalability related to the fact that all data assimilation steps can be performed via memory, also for very large problems. This coupling allows applications for land surface and subsurface models. PDAF was also coupled to oceanic models (Nerger et al., 2006), and more recently atmospheric models.

In contrast to DART and PDAF, which are focused solely on data assimilation, data assimilation utilities also exist within more general model-data informatics systems, such as PEcAn (http://pecanproject.org Dietze et al., 2013; Lebauer et al., 2013). As of writing PEcAn has been coupled to almost 20 ecosystem models, and four of those models (ED2, LINKAGES, SIPNET, and LPJ-GUESS) have implemented the additional modules required to use data assimilation. PEcAn also supports a more generalized ensemble filter based on a multivariate Tobit data model, which is an alternative distribution to the Gaussian and better accommodates zero-bound, proportion, and zero-inflated observations, and a Wishart process error model in place of the traditional assumption that the model's process error is known (Raiho et al., 2020). PEcAn also handles a number of other aspects on top of a data assimilation framework such as job tracking or archiving, input processing, data ingest, and visualization, and can quantify and propagate additional uncertainties, such as parameter uncertainty, parameter variability (random effects), driver uncertainty, and process error.

### 3.6. Computational and Big Data Aspects

The generation of reanalysis products as described in Section 2 and the reanalysis specific algorithmic procedures presented in Sub-Sections 3.1–3.7 are demanding computationally as well as in terms of data volume and data variety. We encourage readers to embrace the technical, computational and big data processing aspects. In the following paragraphs, we review the diverse solutions developed on high performance computers for simulating real world problems and for producing Earth system reanalysis. The examples demonstrate the technical challenges addressed in past, present and future reanalysis.

The evolution of reanalyzes is in line with overall developments in geoscience modeling, which see increases in model's spatial resolution (Prein et al., 2015; Stevens et al., 2019) with simultaneously expanding model domains of regional high-resolution models (Leutwyler et al., 2016), multi-physics fully coupled Earth system models (Eyring et al., 2016; Giorgi & Xue-Jie, 2018; R. R. Shrestha et al., 2014), and many ensemble members in climate change experiments (Eyring et al., 2016) and data assimilation studies (Kurtz et al., 2016; Naz et al., 2020). For example, the 16-years pan-European 3 km soil moisture reanalysis in Naz et al. (2020) with the coupled CLM-PDAF model-data assimilation system and 20 ensemble members accounts for a manageable 100k CPU core hours using 1920 CPU cores simultaneously per data assimilation experiment on a standard Linux cluster with about 7.4 simulated years per day (SYPD) wall clock time. SYPD measures the number of simulated years per 24h wall clock time on any given computational platform. The ERA5 reanalysis (Hersbach et al., 2020) as an example of a global atmospheric reanalysis accounts for a computationally more expensive 0.016–0.025 SYPD. Splitting the simulation time into a number of concurrently running, independent simulation streams makes the computational problem more manageable (Hersbach et al., 2020).

These developments are enabled by steady HPC developments toward massively parallel, heterogeneous (e.g., CPUs and GPUs combined), modular peta-scale supercomputers, providing ever-increasing computational resources (Davis et al., 2012; Dongarra et al., 2018; Schulthess et al., 2019) while making resource efficient simulations on the latest HPC systems technically demanding (Fuhrer et al., 2018). Hence, Earth system models (ESMs) and accompanying software tools have evolved into sophisticated, complex applications (P. Bauer et al., 2015). To make efficient use of today's pre-exascale (GPU-accelerated) HPC systems, performance-portable, highly scalable simulations, processing, analysis and visualization applications and workflows are in constant development (Schulthess et al., 2019). Developments include co-design approaches that combine hardware with software frameworks and algorithmic developments for example, by using domain-specific languages (DSLs; Lawrence et al., 2018), or by increasing memory use efficiency (Fuhrer et al., 2018). This usually requires a refactoring and substantial code modernizations of legacy models (Fuhrer et al., 2018), data processing chains and storage concepts.

Coupled multi-physics Earth system models link the Earth's compartments through mass, energy, momentum transfers and the biogeochemical cycles across multiple spatio-temporal scales and thereby improve the realism of key processes in the Earth system (Eyring et al., 2016; Giorgi & Xue-Jie, 2018; Heinze et al., 2019). Technically, one distinguishes external and internal two-way coupling. Common to all coupling solutions is to strive for modularity, flexibility, and portability of the implementation (Valcke et al., 2012). External couplers such as OASIS3-MCT (Valcke, 2013) or YAC (Hanke et al., 2016), provide a synchronized exchange of information along compartmental interfaces. For external coupling, computations, transformations and communications are usually done concurrently with parallelized codes and in memory (online coupling) to allow scalability of coupled systems. Internal coupling is a more code-intrusive approach which can also be combined with external coupling. Examples of the internal coupling approaches are the Earth System Modeling Framework (ESMF; Collins et al., 2005) or the Modular Earth Submodel System (MESSy; Jockel et al., 2010), that couples unified component submodels that allow for a process based coupling, irrespective of compartmental interfaces (e.g., in atmospheric chemistry). Typical challenges in coupling are load balancing between different component models, which heavily affects the overall performance and thereby also the scaling of the application (Gasper et al., 2014; Valcke, 2013), and the establishment of common software infrastructures with standardized interfaces for purpose-built and generic couplers. Another emerging challenge is the efficient use of modular, heterogeneous supercomputers by multi-component coupled models that, for example, offload certain components on different hardware partitions.

The model and HPC developments, inevitably lead to big data challenges (De Mauro et al., 2016; Kitchin & McArdle, 2016; Overpeck et al., 2011). This is mainly due to the unprecedented data volumes throughout the data lifecycle, affecting data generation, transfers, storage, analysis, dissemination, and archival (Gandomi & Haider, 2015; Overpeck et al., 2011; Schnase et al., 2016). The data volume increase is mainly due to a higher spatio-temporal resolution, shorter output intervals, ensemble runs and more variables. In case of the transition from the global ERA-Interim to the ERA5 reanalysis (See Section 2) horizontal resolution changes from 79 to 31 km, vertical resolution from 60 to 137 levels and temporal resolution from 6- or 3-h to 1-h output intervals with more even variables (Dee et al., 2011; Hersbach et al., 2020), resulting in about 5 petabytes total data volume for ERA5 (Hersbach et al., 2020). Going from ERA-Interim to ERA5 leads, for example, to a factor 80 data volume increase at full resolution in a Lagrangian transport simulation study by (Hoffmann et al., 2019), while there are calls for exascale climate modeling at 1 km global resolution (Palmer, 2016). The simulation, even with massively parallel model systems and parallel data assimilation frameworks, may become memory-limited, depending on the data assimilation algorithm in combination with the ensemble size and resolution (Section 3).

Transitioning toward a full representation of the Earth's water and energy cycles, more input observations from a large number of data sources (radiosonde, aircraft, satellite, soundings, etc.) are used with the generation of a reanalysis, for example, from ERA-40 (S. M. Uppala et al., 2005) to ERA-Interim (Dee et al., 2011), and lately to ERA5 (Hersbach et al., 2020), data variet. also becomes an issue in the data pre-processing for the analysis. Aside from using parallel input and output from and to losslessly compressed, portable data formats, such as netCDF, energy- and time-consuming movements of simulation data can be minimized and the overall data volume reduced by in situ processing, which performs data analysis and visualization in memory (Ayachit et al., 2016; A. C. Bauer et al., 2016; Childs et al., 2019). As this is not always feasible or desirable, big data-capable processing and analysis, that is, doing concurrent computations using shared and distributed memory parallel computing paradigms on heterogeneous HPC systems with generic methods, also become increasingly available (e.g., Krajsek et al., 2018). Finally, as the simulation data has to meet FAIR (Findable, Accessible, Interoperable, and Reusable) principles (M. D. Wilkinson et al., 2016), research data management systems encompass the curation, analysis, publication, re-use, and archival components of the data life cycle. Prominent examples of federated data repositories that provide interoperable (e.g., standardized formats, common meta data conventions, controlled vocabularies, and data reference syntax), provenance-enabled (e.g., unique object identifiers) ECMWF reanalysis and CMIP climate model data in the multi-PB range are the Copernicus Climate Data Store (Buontempo et al., 2020; Hersbach et al., 2020), or the Earth System Grid Federation (ESGF; Cinquini et al., 2014; D. N. Williams et al., 2009), whose infrastructure can be seamlessly integrated via APIs into user workflows.

## 4. Toward Terrestrial Ecosystem Reanalysis

Several important ecosystem processes (photosynthesis, evapotranspiration, surface, or groundwater flow) and variables (carbon fluxes, gross primary production, soil moisture, soil temperature, and terrestrial water storage) are already considered in reanalysis studies of the land surface, hydrological processes or the terrestrial carbon cycle. However, other important aspects (e.g., land use and land cover changes, disturbances, species migration) and variables (genetic composition, species traits, populations, ecosystem structure and functions and others) directly or indirectly related to the aforementioned are not considered in these reanalysis studies. Therefore, we argue to expand terrestrial ecosystem reanalysis beyond the previously discussed fields of application (land surface, hydrology, and terrestrial carbon cycle) by developing a framework that allows to develop a consistent set of products that take into account not only the underlying physical and chemical principles but also biochemical- and biological principles. Terrestrial ecosystem reanalysis will therefore enable better understanding, documentation and assessment of interactions and feedbacks between abiotic and biotic processes at different spatial and temporal scales but also the impacts of disturbances of various origin on the overall functioning of ecosystems and the services they provide. To develop a terrestrial ecosystem reanalysis framework, mathematical- and physical representation of theoretical processes in the form of models informed by both abiotic and biotic data and their interactions and coupling mechanisms are needed. Several recent concepts in ecosystem research provide the foundation for such terrestrial ecosystem reanalysis. These include: (a) integrated ecosystem models considering biotic-abiotic feedback mechanisms, (b) defining a consistent set of Essential Ecosystem Variables (EEV) as other scientific communities have identified them, (c) remote biotic and abiotic trait observations, (d) access to combined biotic and abiotic in-situ observations, and (e) access to specific biotic time series data. In the subsequent sections we briefly outline the elements.

### 4.1. Integrated Ecosystem Models

The challenge to predict and model especially biotic ecosystem variables is not new (Pereira et al., 2013; van den Hoogen et al., 2019). Global dynamic vegetation and land surface models incorporate important biogeophysical processes but lack representation of processes related to biotic variables (Table 2). In line with recent progress (Biber et al., 2020; van den Hoogen et al., 2020) we propose to develop models with stronger abiotic-biotic processes, their interactions and feedback mechanisms to produce meaningful continuous time series which go beyond biologic patterns. Species distribution models (SDMs) are the most common mathematical description in ecological modeling providing spatially explicit predictions based on regressions and/or complex and advanced machine learning models (Deep Learning, Bayes, Regression Trees, and others) in particular on presence-absence and abundance data (Elith & Leathwick, 2009; Jetz et al., 2019; Phillips et al., 2006). SDMs are often supported with predictors. Ecological predictors in the context of SDMs are ecosystem 2–4 D structural (slope, aspect, and altitude) and functional attributes (enhanced vegetation index, land surface temperature, and albedo), several climatic variables and spatially explicit landscape characteristics (Arenas-Castro et al., 2018) which are already output variables of terrestrial reanalysis. Stacked or joint SDMs combine correlative SDMs for multiple species (Biber et al., 2020; D. P. Wilkinson et al., 2019) while SDMs do not provide processes that lead to discovered patterns but infer these processes. There exist ecological process based models focusing on for example, physiology-related mechanisms (Connolly et al., 2017; Kearney & Porter, 2009), simulations on genetic architecture of phenotypes (Schiffers et al., 2014), ecological niche models (Regos et al., 2019), community dynamics, resource competition and biotic interactions (Staniczenko et al., 2017). A combination of all these approaches framed in a macro-ecologic or a macrosystem ecological modeling framework (Cabral et al., 2019; Heffernan et al., 2014; Wuest et al., 2020) currently presents one of the most promising approaches to be used in a global and biotic oriented terrestrial ecosystem reanalysis.

### 4.2. Agreement on Essential Ecosystem Variables

Agreement on key variables of interest is essential for a targeted reanalysis work flow. The conceptual approach of defining essential variables was first applied by the Global Climate Observing System (GCOS). GCOS developed a selection of Essential Climate Variables (ECVs) which critically contribute to the characterization of climate (Bojinski et al., 2014). Marine scientists agreed on a selection of Essential Ocean Variables (EOVs) in 2010 (Lindstrom et al., 2010). The Group on Earth Observations Biodiversity Observation Network (Scholes et al., 2008; GEO BON, Scholes et al. 2008) developed the concept of Essential Biodiversity Variables (EBVs) to monitor changes in biodiversity (Pereira et al., 2013). Although the various Essential Variable (EV) concepts were developed from a disciplinary perspective, they overlap in terms of

10.1029/2020RG000715

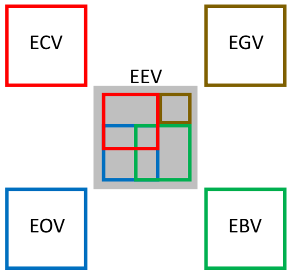


**Figure 4.** Essential Variables were developed for Climate (ECVs), Biodiversity (EBVs), Ocean (EOVs), and Geodiversity (EGVs). Essential Ecosystem Variables (EEVs) for observation and modeling of ecosystems remain to be defined from the perspective of ecosystem reanalysis. Once the EEVs are agreed upon in a brought consensus, the EEVs will be a subset of these four domains bridging the biotic-abiotic variable space which remains a key challenge.

certain variables and their applicability across domains (Figure 4). For example, EBVs and ECVs overlap in the terrestrial domain of ECVs via the *ecosystem functional attributes* and *ecosystem structure*, while the four EBV classes *genetic composition*, *species population*, *species traits*, and *community composition* are beyond the scope of ECVs. Several concepts have been developed that focus on more systemic approaches to derive EV for the description of ecosystem functions. Haase et al. (2018) proposed to merge the concept of EBVs with the ecosystem integrity concept (Muller et al., 2000). Other concepts focusing on the integration of EBVs and ECVs were framed for Essential Ecosystem Variables (Miguel et al., 2017), Essential Ecosystem Functioning Variables (Alcaraz-Segura et al., 2016), and Essential Ecosystem Service Variables (Balvanera et al., 2016).

In line with these efforts, selected elements of the various EV concepts can be combined to EEVs aiming for the balance between abstraction and necessary precision to deliver meaningful answers with regard to monitoring and understanding global change, pressures and disturbance impacts on terrestrial ecosystems. In this sense, EEV can provide a cross-domain information basis for Earth system modeling (Figure 4), but also data interfaces for the coupling of different model and monitoring approaches through utilizing continuous reanalyzed data sets (Figure 3). A reanalysis framework will thus provide better consideration of biotic-abiotic feedbacks that regulate global change at multiple spatial and temporal levels (e.g., disturbance behavior, ecosystem functions, services, and integrity). Further expanding the reanalyzed variables beyond those of Section 2 will be beneficial for quantifying pressures and drivers of for example, biodiversity and geodiversity loss at individual scales (George et al., 2019; Scholes et al., 2008).

Regarding the available measurement methods, a distinction can be made between measurements using (a) campaign-based data collection and inventories (e.g., habitat mapping, soil mapping, and others), (b) in-situ sensors and (c) remote sensing methods. The constantly growing availability of novel sensors and measurement techniques enables new opportunities to combine different measurement methods and approaches and opens up a number of new opportunities to measure and quantify and assess new "essential variables" like efficiencies or ratios, which were previously very difficult or impossible to be measured or derived on relevant scales. The rapid development of remote sensing sensors and techniques on different platforms (wireless sensor networks, drones, air- and space-borne remote sensing) in particular opens up opportunities for monitoring local, continental up to global applications. A remotely sensed and in-situ record coupling approach of an essential variable needs to fulfill a few prerequisites, before being utilized in an ecosystem reanalysis framework:

1. Adequate predictability of the target variable by the observation system
2. Coupling of different spatial, temporal, and ecosystem scales
3. Transparent processing chain and documentation
4. Consistent, trend-preserving time-series
5. Few or no gaps in the record of continuously processed data
6. Adequate accuracy being proven by validation against in situ network data
7. Similar accuracy level for the whole period
8. Possibility of standardization and thus comparability of the data and data products recorded from remote sensing

Those prerequisites are often given for meteorological records, with operational and therefore continuous remote sensing data and data products dating back several decades. However, for terrestrial variables and in-situ data products this operationality is often not given. Long-term essential variable products could be generated by multi-mission records, measured by sensors not built for the purpose and experimental missions lasting for only a few years. Space agencies should consider those issues for the preparation of follow-on missions. Copernicus and GEO/GEOSS missions have this long-term perspective. In the following sections, we provide a few candidate variables to become essential ecosystem variables. While this list is by no means complete, it is intended to instigate further discussion.

### 4.3. Remotely Sensed Trait Observations

Remote sensing of physical and biophysical properties of the Earth System is an essential component for reanalysis of Earth and land surface processes such as energy-related, hydrological and biogeochemical processes (Albergel et al., 2018; Kumar et al., 2019; R. H. Reichle, Draper, et al., 2017). Remotely sensed data provide information to initialize and monitor land surface properties such as land cover (Dee et al., 2011) or are assimilated directly using different methods into the model system such as snow properties (Toure et al., 2018), soil water content (R. Reichle et al., 2016; R. H. Reichle et al., 2019), or leaf area index (Albergel et al., 2018; Kumar et al., 2019). Continuous gapless time series with full spatial coverage are available for hydrological key state variables such as soil moisture (Dorigo et al., 2017), plant properties (plant traits) such as LAI (X. Y. Xie et al., 2019), canopy height and aboveground biomass (Duncanson et al., 2020), gross primary production (Zhang et al., 2017) or plant phenology (Cleland et al., 2007) that are important plant traits for terrestrial ecosystem reanalysis.

In contrast to abiotic states, remotely sensed *biotic* states of ecosystems and specifically biotic trait observations from space (Gamon et al., 2019; Kattge et al., 2020; Lausch et al., 2016, 2019) are still in its infancy. The use of the trait approaches from satellite observations refers to changing the perspective from an ecosystem based point of view to the earth observation based point of view. Remote sensing records traits according to the principles of image spectroscopy over the entire electromagnetic spectrum from the visible to the microwave range (Ustin & Gamon, 2010). These spectral traits are crucial to bridge gaps between in-situ and remote sensing approaches (Lausch et al., 2018).

In the last decade a number of trait concepts were developed which can usefully be incorporated to advance the robust concept of biome specific plant functional types. These trait concepts include plant traits (Kattge et al., 2020) such as leaf traits (Moreno-Martinez et al., 2018) or plant functional traits (Bruelheide et al., 2018), geo-traits (Lausch et al., 2019), or the spectral traits (Deans et al., 2012; Lausch et al., 2016). Plant traits are genetic, anatomical, morphological, biochemical, biophysical, physiological, structural or phenological characteristics, and properties of organisms (Kattge et al., 2020). Traits and their chances (trait-variations) help us to understand, explain and predict where organisms live, how they react to environmental changes and how they interact with different stressors, disorders or resource constraints (Green et al., 2008). "Ecologists are increasingly looking at traits - rather than species - to measure the health of ecosystems" (Cernansky, 2017). Therefore, traits are crucial indicators or filters to measure and assess the ecosystem state, environmental changes, land use intensity, stress or disturbances in ecosystem processes (Averill et al., 2019; Gámez-Virués et al., 2015). Traits exist and can be measured, described and evaluated on all spatio-temporal scales (Abelleira Martínez et al., 2016). They represent a crucial bridging approach to standardization and monitoring of biotic-abiotic interactions (Lausch et al., 2016, 2019).

Global scale data availability meets the reanalysis' demand for high resolution data coverage back in time over decades. The trait observation centered approach has potential to elucidate the largely unknown macrosystem ecological processes (Fei et al., 2016; Gholz & Blood, 2016; Heffernan et al., 2014) and integrate these into ecosystem models. The strong causal connection to earth system processes such as nutrient cycling (Sitzia et al., 2018), surface hydrology (Matheny et al., 2017; Oddershede et al., 2019), productivity of ecosystems (Lees et al., 2018), observing terrestrial ecosystems and the carbon cycle from space (Schimel et al., 2015, 2019), mapping plant functional diversity (Schneider et al., 2017; Stavros et al., 2017) and biodiversity (Averill et al., 2019; Kissling et al., 2018; Lausch et al., 2016) poise the trait observation centered concept for future reanalysis.

### 4.4. In-Situ Long Term Ecosystem Observatories

Data relevant to ecosystem structure and covering a range of biotic-abiotic variables are measured at numerous sites globally. These sites often cover different ecosystem compartments providing ground truthing to satellite data and important information to study multi-scale processes from site-to-global scale. Scientific networks undertaking these measurements include the Critical Zone Observatories (Brantley et al., 2017), the Integrated Carbon Observation System (ICOS, Lavric et al., 2016) and FluxNet network (Papale et al., 2006), and the continental-global Long Term Ecosystem Research Networks (LTER, Keller et al., 2008; Mirtl et al., 2018; Teeri & Raven, 2002). The Critical Zone Observatories focus on geophysical characterizations of the zone between bedrock and atmosphere (Baatz et al., 2018; Weintraub et al., 2019). The observation strategy of CZO sites is driven by the site's principle investigators. This means that observation technology, observation frequency, and data reporting are largely not congruent for many sites, making integration of data into models and characterizing observation uncertainty difficult over large areas. Recent efforts incentivize more standardized observation technology and data reporting (Brantley et al., 2017). The operational Integrated Carbon Observation System (Lavric et al., 2016) provides continuous, high quality and standardized abiotic and biotic measurements through a dedicated network of observation stations in Europe (https://www.icos-cp.eu/). The main goal of ICOS is the detection and quantification of European greenhouse gas fluxes. Protocols standardize target observations, pre-scribe transparent data processing for sites, for a consistent level of accuracy for the observation period. Notable biotic-abiotic ICOS observations in the standard protocol for level 1 sites are above- and below-ground biomass, net ecosystem exchange, CO2, H2O, H, CH4, and N2O exchange between land surface and atmosphere, and more. ICOS sites therefore leverage data quality of European FluxNet sites to enable calibration, validation and data assimilation into carbon- and ecosystem reanalysis.

LTER enables ecosystem research and observation since decades (Mirtl et al., 2018). LTER focusses on ecosystem research and observations of ecosystem relevant biotic, ecologic variables. In Europe, the emerging Integrated European Long-Term Ecosystem, critical zone & socio-ecological Research Infrastructure (eLTER RI, http://www.lter-europe.net/elter-esfri) aims at the systematic, harmonized and sustainable gathering of integrated terrestrial ecosystem data across 250 sites to enable ecosystem reanalysis at the European scale. The terrestrial ecosystem reanalysis strategy was also addressed in the foundation of the National Ecological Observatory Network (NEON) of the United States (Keller et al., 2008; Teeri & Raven, 2002). Monitoring approaches of eLTER RI and NEON aim at the coverage of bio-ecological, geoscientific, and socio-ecological components as well as representative coverage of the most important ecological, climatic and geographical gradients. Besides site level observations, eLTER RI fosters integration of long-term monitoring schemes with the United Nations Economic Commission for Europe (UNECE) Working Group International Cooperative Program (ICP) on "Integrated Monitoring of Ecosystems" and system-focused sites of ICP Forest. Given the implementation of these standardized protocols and data reporting for multi-variate in-situ observations (Firbank et al., 2017), the networks provide a database for site-level to continental scale terrestrial ecosystem reanalysis as has been demonstrated already for FluxNet and river discharge monitoring in Section 2.

### 4.5. Existing Specific Biotic In-Situ Data

Biotic data can also be rather abstract conceptual data such as species threat indices of abundance estimated from population samples, or it can be detailed phylogenetic, taxonomic and family specific data. Proenca et al. (2017) list monitoring schemes including animal and bird life of international character with at least continental coverage reaching back to 1966. While in those early years, only few countries applied national monitoring programs, by 2005 already 20 European countries adopted standardized bird-monitoring programs (Gregory et al., 2005), and by 2010, the efforts reached global scale through the Earth Observation Biodiversity Network embracing less developed regions (GEO BON, Pereira et al., 2013). The biodiversity data are an indicator for changes in land management, land use and their intensity, ecosystem structure, its functionality, state and parameterization (e.g., Kamp et al., 2018; Pereira et al., 2013). While biodiversity is often stronger affected by land use, fungi and fungal groups are impacted stronger by climatic drivers and soil states such as calcium and phosphorus concentration, and pH (Tedersoo et al., 2014). The Projecting Response of Ecological Diversity In Changing Terrestrial Systems (PREDICTS, Hudson et al., 2017) is a database resulting from anterior efforts, characterizing numerous aspects of biotic populations in a structured way. One data source mined by PREDICTS is the Global Biodiversity Information Facility (GBIF, 2020, https://www.gbif.org). GBIF offers more than 1 billion records within 42,000 data sets on biodiversity and trait information in a citable and open source manner while unstructured. One particularly large data set on flora and fauna characteristics is the Plant Traits Database with more than 6.9 million trait records and 148,000 plant taxa (e.g., Kattge et al., 2011, 2020). The international tree ring data bank is another valuable set of spatio-temporal data to include useful biogeographical information through dendrochronologic measurements (Babst et al., 2017). Although challenges remain with regard to spatial, environmental, and taxonomic representativity (S. Zhao et al., 2019), dendrochronological data are an important indicator for ecosystem stress through heat, drought, and flooding (A. P. Williams et al., 2020) frequently used in climate reconstruction over centuries.

In addition to these diverse but networked long term in situ data, high-throughput Biodiversity Data (HBD, Makiola et al., 2020; Wuest et al., 2020) are becoming available in the near future. HBD on environmental DNA will be available from DNA and RNA sequencing at selected sites (Bush et al., 2017). Also called next-generation biomonitoring, HBD produces data sets on ecosystem compositions and allows macroecological modeling at species to community level (Bani et al., 2020). Previously mentioned modeling frameworks jointly with HBD may allow the resolution of species specific causalities for macrosystem-ecological modeling (Cabral et al., 2019; Wuest et al., 2020) to establish a robust data foundation for producing reanalyzed biotic ecosystem states. Which data can be assimilated in a reanalysis or model data fusion framework remains a question of harmonization and standardization for the specific reanalysis objective.

### 4.6. Example of a Site Level Ecosystem Reanalysis for the Harvard Forest

Raiho et al. 2020 provides an example of assimilating dendrochronological data into a forest gap model (H. H. Shugart & Smith, 1996) to understand what drives uncertainty in hindcasts of forest stand development.

Forest stand development is driven by competition between tree species for light, water, and nutrients initializing when a gap is made in the forest canopy by a disturbance. Following canopy disturbance, initially fast-growing shade intolerant species are dominant and eventually slow-growing, shade tolerant species become dominant (H. H. Shugart, Jr. & West, 1980). The forest gap model, LINKAGES, simulates competition between individual trees by incrementing each individual tree's diameter at breast height (DBH) based on limiting environmental factors: available light, soil moisture, and soil nitrogen (Post & Pastor, 1996; Figure 5a). These DBH values are then aggregated into annual species-level aboveground biomass using allometric relationships for each tree species. These species-level aboveground biomass measurements are then easily comparable to both terrestrial models that do not produce individual level DBH values and a variety of ecosystem data sources including tree ring derived aboveground biomass measurements. Typically, forest gap models are run without any data constraints on aboveground biomass. Raiho et al. 2020 use data assimilation to constrain LINKAGES biomass reconstructions with dendrochronological data, which improves model simulation of competition between individual trees. To assimilate species-level biomass into LINKAGES, Raiho et al. 2020 develop the Tobit Wishart Ensemble Filter (TWEnF). The TWEnF works very similarly to an EnKF but accounts for zero-truncated data, common in ecological datasets, and incorporates an estimate of process error (Raiho et al., 2020).

To demonstrate the ability of the TWEnF to improve species biomass reconstruction, Raiho et al. 2020 consider the well-documented forest site, the Lyford Plot at Harvard Forest (Eisen, 2015). Land at Harvard Forest was cleared for agriculture (Foster et al., 1998) in 1900. The Lyford plot spans 3 ha in a temperate climate and thin glacial till soil type. Since 1900, the Lyford plot has been regrowing temperate deciduous trees including: red oak (*Quercus rubra*), red maple (*Acer rubrum*), yellow birch (*Betula alleghaniensis*), American beech (*Fagus granifolia*), and eastern hemlock (*Tsuga canadensis*). Within the Lyford plot red oak has become the dominant species (Figure 5b), unlike the surrounding forests where red maple has become the dominant (Abrams, 1998). We present an analysis at one site, noting this type of analysis could be expanded to scales where large scale tree ring data are available (See international tree ring database, ITRDB, Babst et al., 2017).

We show results for an open loop run without data assimilation, constrained model predictions with partial data assimilation, and model predictions with data assimilation of the full time period (Figure 5c). For simplicity in this example, we show total aboveground biomass which is the sum of species-level biomasses. Figure 5c demonstrates how the reconstruction of total aboveground biomass of red oak (*Quercus rubra*) at Harvard Forest changes depending on the amount of data that was assimilated. In the open loop scenario, there is very large uncertainty about the total aboveground biomass in 2010 because we incorporated parameter, meteorological, and process uncertainties. This depicts a realistic forecast including uncertainty of over 50 years of forest stand development. Raiho et al. (2020) found that constraining initial conditions alone with data also had a large effect on reconstruction uncertainty. The second partial data assimilation constraint shows this effect where assimilating data to 1984 greatly constrains uncertainty in 2010 but does

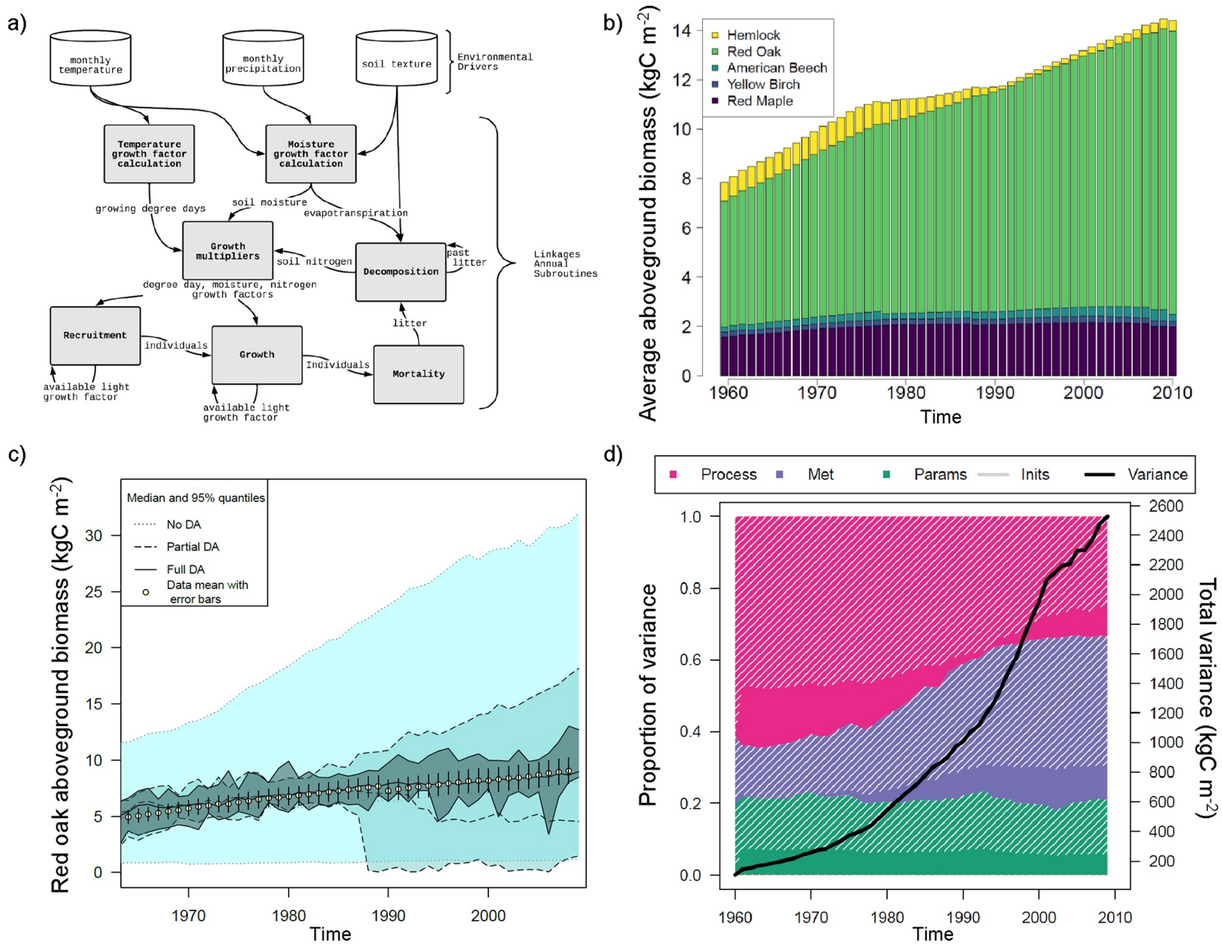


**Figure 5.** (a) Conceptual diagram of the forest gap model LINKAGES. Each top cylinder represents the inputs to the model, and gray boxes represent a subroutine within the model. (b) The stacked bar plot shows the posterior mean of species-level estimated biomass from tree rings collected at the Harvard Forest Lyford Plot. (c) Total aboveground biomass distribution for the dominant species red oak (*Quercus rubra*) simulated by three LINKAGES runs are shown in this time series with different amounts of data constraint, demonstrating the usefulness of reanalysis on improving predictions from LINKAGES. (d) The relative contribution of each type of variance to total aboveground biomass variance. Hashed areas are the relative variances that can be attributed to the covariance with initial conditions. Over time initial condition uncertainty covariance together with meteorological uncertainty (purple) accounted for a larger proportion of total variance while initially process uncertainty dominates. Total variance over time is shown as black line with the scale on the right hand *y*-axis.

not result in red oak dominance. Interestingly, this is similar to a prediction made by Lorimer (1984). Finally, the full data assimilation scenario shows that the TWEnF constrains aboveground biomass predictions well. Furthermore, the relative contribution of each type of variance to total aboveground biomass variance was analyzed. Individual contributions change over the assimilation period, providing information on reasons and origins of forecast uncertainty.

The site-specific reanalysis of species-level aboveground biomass demonstrates how data assimilation can be used to provide further insights into processes represented in models and improve predictions. Comparing the partial data assimilation to the full data assimilation scenario demonstrates the importance of iteratively updating model forecasts using new observations. The results show the framework's ability to predict competitive dynamics between forest species counterintuitive to conservative experts and model prediction. The framework enabled the model to learn the competitive mechanisms behind red oak outperforming red maple populations under sites-specific conditions with fewer canopy disturbances. Furthermore, variance partitioning between the uncertainty components, parameter, meteorological, process, and initial conditions, revealed which aspects of uncertainty were most important to constrain to improve long-term prediction (Figure 5d). Initial conditions were found to be most important over the time period, where including initial conditions reduced total forecast uncertainty over a 50-years time period by more than 20%. Process uncertainty was also an important aspect of uncertainty initially followed by an increase in meteorological uncertainty over time.

## 5. Conclusions

The review analyzes the state-of-the-art in methods, recent developments and prospects of reanalysis for three subcomponents of the Earth system (atmosphere, ocean, and land). It focusses on 21st century reanalysis products including ecosystem reanalyzes. Other studies reviewed reanalysis in climate and atmospheric science (Dee et al., 2011; Fujiwara et al., 2017), and the physical (Storto et al., 2017) and biogeochemical oceanic research (Fennel et al., 2019). We outline major advances using exemplary reanalysis products while we acknowledge that the discussion of several excellent reanalysis products is not included here for reason of space. Recently, networks of distributed in-situ sensors such as buoys and biogeochemical Argo floats (Fennel et al., 2019), eddy covariance stations (Peylin et al., 2016), surface water runoff observations (Hobeichi et al., 2019) and meteorological station data (Hersbach et al., 2020) were used in reanalysis of physical and biogeochemical Earth system processes. These reanalyzes highlight progress in predicting particularly terrestrial ECV. Novel concepts on global biotic trait observation approaches (e.g., plant traits, TRY trait data base, Kattge et al., 2011, 2020), coupled with satellite based earth observation for spectral traits (Lausch et al., 2019), and integrated ecosystem observatories at continental (e.g., Europe: eLTER Mirtl et al., 2018) to global scale (GEO BON, Pereira et al., 2013) promise meaningful harmonized and standardized in-situ data including more biotic variables for reanalysis. This allows for establishing new variable concepts, merging ECVs and EBVs into EEVs. In this sense, a prioritization of environmental variables against the background of the data requirements of the reanalysis can also provide important support for the development of future integrated environmental monitoring programs. Specific and targeted new essential variables are set forth to address the needed terrestrial biotic-abiotic feedback mechanisms at a hyper-resolution global modeling scale through terrestrial ecosystem reanalysis.

**Acknowledgments**
R. Baatz, C. Poppe, and M. Mirtl acknowledge funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871128 (eLTER PLUS). H. J. Hendricks Franssen, C. Montzka, and H. Vereecken acknowledge support from the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy, EXC-2070–390732324–PhenoRob. E. Euskirchen acknowledges the Biological and Environmental Research (BER) program within the U.S. Department of Energy (DOE) as part of the Next-Generation Ecosystem Experiments in the Arctic (NGEE Arctic) project as well as DOE grant No. DE-SC0012704. D. Sihi acknowledges support from Emory University's Halle Institute for Global Research and the Halle Foundation Collaborative Research Grant. MCD was supported by NSF 1638577, NSF 1458021, and NASA 16-CMS16-0007. S. Ciavatta was supported by the Natural Environment Research Council (NERC) NCEO and by the NERC CAMPUS project, as well as by the SEAMLESS project, which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004032. K. Fennel acknowledges support from the NSERC Discovery program, the Ocean Frontier Institute, and MEOPAR. G. De Lannoy was supported by the Research Foundation Flanders (FWO) G0A7320N and the ESA CCN1 CCI + Soil Moisture Scientific Evolution project. U, Mishra was supported through a grant from U.S. Department of Energy under Sandia National Laboratories contract DE-NA-0003525. L. Samaniego acknowledges the Advanced Earth System Modeling Capacity project funded by the Helmholtz Association. S Zacharias and H. Beck acknowledge the support of the TERENO (Terrestrial Environmental Observatories) funded by the Helmholtz Association. A. Fox acknowledges funding from NASA Terrestrial Ecosystems Grant 80NSSC19M0103.
Open access funding enabled and organized by Projekt DEAL.

## Data Availability Statement

No data were used in producing this manuscript.

## References

Abelleira Martínez, O. J., Fremier, A. K., Günter, S., Ramos Bendaña, Z., Vierling, L., Galbraith, S. M., et al. (2016). Scaling up functional traits for ecosystem services with remote sensing: Concepts and methods. *Ecology and Evolution*, *6*(13), 4359–4371. https://doi.org/10.1002/ece3.2201

Abrams, M. D. (1998). The red maple paradox: What explains the widespread expansion of red maple in eastern forests? *BioScience*, *48*(5), 355–364. https://doi.org/10.2307/1313374

Ainsworth, E. A., & Long, S. P. (2005). What have we learned from 15 years of free-air $CO_2$ enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising $CO_2$. *New Phytologist*, *165*(2), 351–372. https://doi.org/10.1111/j.1469-8137.2004.01224.x

Albergel, C., Dorigo, W., Reichle, R. H., Balsamo, G., de Rosnay, P., Munoz-Sabater, J., et al. (2013). Skill and global trend analysis of soil moisture from reanalyses and microwave remote sensing. *Journal of Hydrometeorology*, *14*(4), 1259–1277. https://doi.org/10.1175/jhm-d-12-0161.1

Albergel, C., Munier, S., Bocher, A., Bonan, B., Zheng, Y. J., Draper, C., et al. (2018). LDAS-Monde sequential assimilation of satellite derived observations applied to the contiguous US: An ERA-5 driven reanalysis of the land surface variables. *Remote Sensing*, *10*(10). https://doi.org/10.3390/rs10101627

Alcaraz-Segura, D., Bagnato, C. E., Paruelo, J. M., Berbery, E. H., Cabello, J., Castro, A., et al. (2016). *Remote sensing of essential ecosystem functional variables*. American Geophysical Union. Retrieved from https://ui.adsabs.harvard.edu/abs/2016AGUFM.B51I.01A

Allen, J. I., Eknes, M., & Evensen, G. (2003). An ensemble Kalman filter with a complex marine ecosystem model: Hindcasting phytoplankton in the Cretan Sea. *Annales Geophysicae*, *21*(1), 399–411. https://doi.org/10.5194/angeo-21-399-2003

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., & Avellano, A. (2009). The data assimilation research testbed a community facility. *Bulletin of the American Meteorological Society*, *90*(9), 1283–1296. https://doi.org/10.1175/2009bams2618.1

Andersson, E., & Jarvinen, H. (1999). Variational quality control. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 697–722. https://doi.org/10.1002/qj.497.v125:556

Arenas-Castro, S., Goncalves, J., Alves, P., Alcaraz-Segura, D., & Honrado, J. P. (2018). Assessing the multi-scale predictive ability of ecosystem functional attributes for species distribution modelling. *PloS One*, *13*(6). https://doi.org/10.1371/journal.pone.0199292

Averill, C., Bhatnagar, J. M., Dietze, M. C., Pearse, W. D., & Kivlin, S. N. (2019). Global imprint of mycorrhizal fungi on whole-plant nutrient economics. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(46), 23163–23168. https://doi.org/10.1073/pnas.1906655116

Axell, L., & Liu, Y. (2016). Application of 3-D ensemble variational data assimilation to a Baltic Sea reanalysis 1989-2013. In *Tellus Series a-Dynamic Meteorology and Oceanography* (Vol. 68). https://doi.org/10.3402/tellusa.v68.24220

Ayachit, U., Whitlock, B., Wolf, M., Loring, B., Geveci, B., Lonie, D., & Bethel, E. W. (2016). The SENSEI generic in situ interface. In *2016 second workshop on in situ infrastructures for enabling extreme-scale Analysis and visualization (ISAV)*.

Baatz, R., Sullivan, P. L., Li, L., Weintraub, S. R., Loescher, H. W., Mirtl, M., et al. (2018). Steering operational synergies in terrestrial observation networks: Opportunity for advancing Earth system dynamics modelling. *Earth System Dynamics*, *9*(2), 593–609. https://doi.org/10.5194/esd-9-593-2018

Babst, F., Poulter, B., Bodesheim, P., Mahecha, M. D., & Frank, D. C. (2017). Improved tree-ring archives will support earth-system science. *Nature Ecology & Evolution*, *1*(2), 0008. https://doi.org/10.1038/s41559-016-0008

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., et al. (2001). FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, *82*(11), 2415–2434. https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2

Balmaseda, M. A., Hernandez, F., Storto, A., Palmer, M. D., Alves, O., Shi, L., et al. (2015). The ocean reanalyses intercomparison project (ORA-IP). *Journal of Operational Oceanography*, *8*, S80–S97. https://doi.org/10.1080/1755876x.2015.1022329

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., et al. (2015). ERA-Interim/Land: A global land surface reanalysis data set. *Hydrology and Earth System Sciences*, *19*(1), 389–407. https://doi.org/10.5194/hess-19-389-2015

Balsamo, G., Viterbo, P., Beljaars, A., van den Hurk, B., Hirschi, M., Betts, A. K., & Scipal, K. (2009). A revised hydrology for the ECMWF Model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *Journal of Hydrometeorology*, *10*(3), 623–643. https://doi.org/10.1175/2008jhm1068.1

Balvanera, P., Cord, A. F., deClerck, F., Drakou, E. G., Geijzendorffer, I., Geller, G. N., et al. (2016). Essential Ecosystem Service Variables. In *GEOBON Open Science Conference*. Retrieved from https://research.utwente.nl/en/publications/essential-ecosystem-service-variables(5e42d97d-682f-4189-aab2-f2fa11b88c6a).html

Bani, A., De Brauwer, M., Creer, S., Dumbrell, A. J., Limmon, G., Jompa, J., et al. (2020). Chapter ten - informing marine spatial planning decisions with environmental DNA. In A. J. Dumbrell, E. C. Turner, & T. M. Fayle (Eds.), *Advances in ecological research* (Vol. 62, pp. 375–407). Academic Press. https://doi.org/10.1016/bs.aecr.2020.01.011

Bauer, A. C., Abbasi, H., Ahrens, J., Childs, H., Geveci, B., Klasky, S., et al. (2016). In situ methods, infrastructures, and applications on high performance computing platforms. *Computer Graphics Forum*, *35*(3), 577–597. https://doi.org/10.1111/cgf.12930

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*(7567), 47–55. https://doi.org/10.1038/nature14956

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I. J. M., et al. (2019). MSWEP V2 Global 3-Hourly 0.1 degrees precipitation: Methodology and Qquantitative assessment. *Bulletin of the American Meteorological Society*, *100*(3), 473–500. https://doi.org/10.1175/bams-d-17-0138.1

Bengtsson, L., Kanamitsu, M., Kallberg, P., & Uppala, S. (1982). Fgge 4-dimensional data assimilation at Ecmwf. *Bulletin of the American Meteorological Society*, *63*(1), 29–43. https://doi.org/10.1175/1520-0477-63.3.277

Bengtsson, L., & Shukla, J. (1988). Integration of space and insitu observations to study global climate change. *Bulletin of the American Meteorological Society*, *69*(10), 1130–1143. https://doi.org/10.1175/1520-0477(1988)069<1130:iosais>2.0.co;2

Bennett, A. F., & Budgell, W. P. (1987). Ocean data assimilation and the Kalman filter - spatial regularity. *Journal of Physical Oceanography*, *17*(10), 1583–1601. https://doi.org/10.1175/1520-0485(1987)017<1583:odaatm>2.0.co;2

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Menard, C. B., et al. (2011). The Joint UK Land Environment Simulator (JULES), model description - Part 1: Energy and water fluxes. *Geoscientific Model Development*, *4*(3), 677–699. https://doi.org/10.5194/gmd-4-677-2011

Biber, M. F., Voskam, A., Niamir, A., Hickler, T., & Hof, C. (2020). A comparison of macroecological and stacked species distribution models to predict future global terrestrial vertebrate richness. *Journal of Biogeography*, *47*(1), 114–129. https://doi.org/10.1111/jbi.13696

Bloom, A. A., Exbrayat, J. F., van der Velde, I. R., Feng, L., & Williams, M. (2016). The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(5), 1285–1290. https://doi.org/10.1073/pnas.1515160113

Bloom, A. A., & Williams, M. (2015). Constraining ecosystem carbon dynamics in a data-limited world: Integrating ecological "common sense" in a model-data fusion framework. *Biogeosciences*, *12*(5), 1299–1315. https://doi.org/10.5194/bg-12-1299-2015

Bocquet, M., & Sakov, P. (2014). An iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, *140*(682), 1521–1535. https://doi.org/10.1002/qj.2236

Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., & Zemp, M. (2014). The concept of essential climate variables in support of climate research, applications, and policy. *Bulletin of the American Meteorological Society*, *95*(9), 1431–1443. https://doi.org/10.1175/bams-d-13-00047.1

Bollmeyer, C., Keller, J. D., Ohlwein, C., Wahl, S., Crewell, S., Friederichs, P., et al. (2015). Towards a high-resolution regional reanalysis for the European CORDEX domain. *Quarterly Journal of the Royal Meteorological Society*, *141*(686), 1–15. https://doi.org/10.1002/qj.2486

Bonan, G. B., & Doney, S. C. (2018). Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. *Science*, *359*(6375), 533. https://doi.org/10.1126/science.aam8328

Bonan, G. B., Levis, S., Sitch, S., Vertenstein, M., & Oleson, K. W. (2003). A dynamic global vegetation model for use with climate models: Concepts and description of simulated vegetation dynamics. *Global Change Biology*, *9*(11), 1543–1566. https://doi.org/10.1046/j.1365-2486.2003.00681.x

Bosilovich, M. G., Chen, J. Y., Robertson, F. R., & Adler, R. F. (2008). Evaluation of global precipitation in reanalyses. *Journal of Applied Meteorology and Climatology*, *47*(9), 2279–2299. https://doi.org/10.1175/2008jamc1921.1

Boussetta, S., Balsamo, G., Beljaars, A., Kral, T., & Jarlan, L. (2013). Impact of a satellite-derived leaf area index monthly climatology in a global numerical weather prediction model. *International Journal of Remote Sensing*, *34*(9–10), 3520–3542. https://doi.org/10.1080/01431161.2012.716543

Brantley, S. L., McDowell, W. H., Dietrich, W. E., White, T. S., Kumar, P., Anderson, S. P., et al. (2017). Designing a network of critical zone observatories to explore the living skin of the terrestrial Earth. *Earth Surface Dynamics*, *5*(4), 841–860. https://doi.org/10.5194/esurf-5-841-2017

Bromwich, D. H., Wilson, A. B., Bai, L., Liu, Z., Barlage, M., Shih, C. F., et al. (2018). The arctic system reanalysis, version 2. *Bulletin of the American Meteorological Society*, *99*(4), 805–828. https://doi.org/10.1175/bams-d-16-0215.1

Bruelheide, H., Dengler, J., Purschke, O., Lenoir, J., Jimenez-Alfaro, B., Hennekens, S. M., et al. (2018). Global trait-environment relationships of plant communities. *Nature Ecology & Evolution*, *2*(12), 1906–1917.

Buontempo, C., Hutjes, R., Beavis, P., Berckmans, J., Cagnazzo, C., Vamborg, F., et al. (2020). Fostering the development of climate services through copernicus climate change service (C3S) for agriculture applications. *Weather and Climate Extremes*, *27*, 100226. https://doi.org/10.1016/j.wace.2019.100226

Burgers, G., Jan van Leeuwen, P., & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, *126*(6), 1719–1724. https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2

Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., et al. (2017). Connecting Earth observation to high-throughput biodiversity data. *Nature Ecology & Evolution*, *1*(7), 0176. https://doi.org/10.1038/s41559-017-0176

Cabral, J. S., Wiegand, K., & Kreft, H. (2019). Interactions between ecological, evolutionary and environmental processes unveil complex dynamics of insular plant diversity. *Journal of Biogeography*, *46*(7), 1582–1597. https://doi.org/10.1111/jbi.13606

Candiani, G., Carnevale, C., Finzi, G., Pisoni, E., & Volta, M. (2013). A comparison of reanalysis techniques: Applying optimal interpolation and ensemble Kalman filtering to improve air quality monitoring at mesoscale. *The Science of the Total Environment*, *458*, 7–14. https://doi.org/10.1016/j.scitotenv.2013.03.089

Carrera, J., & Neuman, S. P. (1986). Estimation of aquifer parameters under transient and steady state conditions: 2. Uniqueness, stability, and solution algorithms. *Water Resources Research*, *22*(2), 211–227. https://doi.org/10.1029/wr022i002p00211

Carrera, M. L., Belair, S., & Bilodeau, B. (2015). The Canadian land data assimilation system (CaLDAS): Description and synthetic evaluation study. *Journal of Hydrometeorology*, *16*(3), 1293–1314. https://doi.org/10.1175/jhm-d-14-0089.1

Carton, J. A., Chepurin, G. A., & Chen, L. G. (2018). SODA3: A new ocean climate reanalysis. *Journal of Climate*, *31*(17), 6967–6983. https://doi.org/10.1175/jcli-d-18-0149.1

Castro-Morales, K., Schurmann, G., Kostler, C., Rodenbeck, C., Heimann, M., & Zaehle, S. (2019). Three decades of simulated global terrestrial carbon fluxes from a data assimilation system confronted with different periods of observations. *Biogeosciences*, *16*(15), 3009–3032. https://doi.org/10.5194/bg-16-3009-2019

Cernansky, R. (2017). Biodiversity moves beyond counting species. *Nature*, *546*(7656), 22–24. https://doi.org/10.1038/546022a

Chabbi, A., LOESCHER, H. W., Tye, M. R., & Hudnut, D. (2017). *Integrated experimental research infrastructures: A paradigm shift to face an uncertain world and innovate for societal benefit. Terrestrial ecosystem research infrastructures* (p. 534). CRC Press.

Chen, B. Y., & Liu, Z. Z. (2016). Global water vapor variability and trend from the latest 36year (1979 to 2014) data of ECMWF and NCEP reanalyses, radiosonde, GPS, and microwave satellite. *Journal of Geophysical Research: Atmospheres*, *121*(19), 11442–11462. https://doi.org/10.1002/2016jd024917

Chen, Y., & Zhang, D. X. (2006). Data assimilation for transient flow in geologic formations via ensemble Kalman filter. *Advances in Water Resources*, *29*(8), 1107–1122. https://doi.org/10.1016/j.advwatres.2005.09.007

Childs, H., Bennett, J., Garth, C., Hentschel, B., & Rhyne, T.-M. (2019). In situ visualization for computational science. *IEEE Computer Graphics and Applications*, *39*(6), 76–85. https://doi.org/10.1109/mcg.2019.2936674

Ciavatta, S., Brewin, R. J. W., Skakala, J., Polimene, L., de Mora, L., Artioli, Y., & Allen, J. I. (2018). Assimilation of ocean-color plankton functional types to improve marine ecosystem simulations. *Journal of Geophysical Research: Oceans*, *123*(2), 834–854. https://doi.org/10.1002/2017jc013490

Ciavatta, S., Kay, S., Brewin, R. J. W., Cox, R., Di Cicco, A., Nencioli, F., et al. (2019). Ecoregions in the Mediterranean Sea through the reanalysis of phytoplankton functional types and carbon fluxes. *Journal of Geophysical Research: Oceans*, *124*(10), 6737–6759. https://doi.org/10.1029/2019jc015128

Ciavatta, S., Kay, S., Saux-Picart, S., Butenschon, M., & Allen, J. I. (2016). Decadal reanalysis of biogeochemical indicators and fluxes in the North West European shelf-sea ecosystem. *Journal of Geophysical Research: Oceans*, *121*(3), 1824–1845. https://doi.org/10.1002/2015jc011496

Ciavatta, S., Torres, R., Martinez-Vicente, V., Smyth, T., Dall'Olmo, G., Polimene, L., & Allen, J. I. (2014). Assimilation of remotely-sensed optical properties to improve marine biogeochemistry modelling. *Progress in Oceanography*, *127*, 74–95. https://doi.org/10.1016/j.pocean.2014.06.002

Cinquini, L., Crichton, D., Mattmann, C., Harney, J., Shipman, G., Wang, F. Y., et al. (2014). The Earth system grid federation: An open infrastructure for access to distributed geospatial data. *Future Generation Computer Systems-the International Journal of Grid Computing and Escience*, *36*, 400–417. https://doi.org/10.1016/j.future.2013.07.002

Clark, E. A., Sheffield, J., van Vliet, M. T. H., Nijssen, B., & Lettenmaier, D. P. (2015). Continental runoff into the oceans (1950-2008). *Journal of Hydrometeorology*, *16*(4), 1502–1520. https://doi.org/10.1175/jhm-d-14-0183.1

Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., et al. (2001). Ecological forecasts: An emerging imperative. *Science*, *293*(5530), 657–660. https://doi.org/10.1126/science.293.5530.657

Cleland, E. E., Chuine, I., Menzel, A., Mooney, H. A., & Schwartz, M. D. (2007). Shifting plant phenology in response to global change. *Trends in Ecology & Evolution*, *22*(7), 357–365. https://doi.org/10.1016/j.tree.2007.04.003

Collins, N., Theurich, G., DeLuca, C., Suarez, M., Trayanov, A., Balaji, V., et al. (2005). Design and implementation of components in the Earth system modeling framework. *International Journal of High Performance Computing Applications*, *19*(3), 341–350. https://doi.org/10.1177/1094342005056120

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., et al. (2011). The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, *137*(654), 1–28. https://doi.org/10.1002/qj.776

Connolly, S. R., Keith, S. A., Colwell, R. K., & Rahbek, C. (2017). Process, mechanism, and modeling in macroecology. *Trends in Ecology & Evolution*, *32*(11), 835–844. https://doi.org/10.1016/j.tree.2017.08.011

Cossarini, G., Mariotti, L., Feudale, L., Mignot, A., Salon, S., Taillandier, V., et al. (2019). Towards operational 3D-Var assimilation of chlorophyll biogeochemical-argo float data into a biogeochemical model of the Mediterranean Sea. *Ocean Modelling*, *133*, 112–128. https://doi.org/10.1016/j.ocemod.2018.11.005

Costanza, R., dArge, R., deGroot, R., Farber, S., Grasso, M., Hannon, B., et al. (1997). The value of the world's ecosystem services and natural capital. *Nature*, *387*(6630), 253–260. https://doi.org/10.1038/387253a0

Dalmonech, D., & Zaehle, S. (2013). Towards a more objective evaluation of modelled land-carbon trends using atmospheric $CO_2$ and satellite-based vegetation activity observations. *Biogeosciences*, *10*(6), 4189–4210. https://doi.org/10.5194/bg-10-4189-2013

Damm, A., Elbers, J., Erler, A., Gioli, B., Hamdi, K., Hutjes, R., et al. (2010). Remote sensing of sun-induced fluorescence to improve modeling of diurnal courses of gross primary production (GPP). *Global Change Biology*, *16*(1), 171–186. https://doi.org/10.1111/j.1365-2486.2009.01908.x

Davis, N. E., Robey, R., Ferenbaugh, C. R., Nicholaeff, D., & Trujillo, D. P. (2012). Paradigmatic shifts for exascale supercomputing. *The Journal of Supercomputing*, *62*(2), 1023–1044. https://doi.org/10.1007/s11227-012-0789-3

Deans, A. R., Yoder, M. J., & Balhoff, J. P. (2012). Time to change how we describe biodiversity. *Trends in Ecology & Evolution*, *27*(2), 78–84. https://doi.org/10.1016/j.tree.2011.11.007

Decharme, B., & Douville, H. (2006). Introduction of a sub-grid hydrology in the ISBA land surface model. *Climate Dynamics*, *26*(1), 65–78. https://doi.org/10.1007/s00382-005-0059-7

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597.

De Lannoy, G. J. M., & Reichle, R. H. (2016). Global assimilation of multiangle and multipolarization SMOS brightness temperature observations into the GEOS-5 catchment land surface model for soil moisture estimation. *Journal of Hydrometeorology*, *17*(2), 669–691. https://doi.org/10.1175/jhm-d-15-0037.1

De Lannoy, G. J. M., Reichle, R. H., Houser, P. R., Arsenault, K. R., Verhoest, N. E. C., & Pauwels, V. R. N. (2010). Satellite-scale snow water equivalent assimilation into a high-resolution land surface model. *Journal of Hydrometeorology*, *11*(2), 352–369. https://doi.org/10.1175/2009jhm1192.1

De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of big data based on its essential features. *Library Review*, *65*(3), 122–135. https://doi.org/10.1108/LR-06-2015-0061

Deng, Y. H., Wang, S. J., Bai, X. Y., Luo, G. J., Wu, L. H., Chen, F., et al. (2020). Spatiotemporal dynamics of soil moisture in the karst areas of China based on reanalysis and observations data. *Journal of Hydrology*, *585*, 124744. https://doi.org/10.1016/j.jhydrol.2020.124744

de Rosnay, P., Drusch, M., Vasiljevic, D., Balsamo, G., Albergel, C., & Isaksen, L. (2013). A simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF. *Quarterly Journal of the Royal Meteorological Society*, *139*(674), 1199–1213. https://doi.org/10.1002/qj.2023

Dietze, M. C. (2017). Prediction in ecology: A first-principles framework. *Ecological Applications*, *27*(7), 2048–2060. https://doi.org/10.1002/eap.1589

Dietze, M. C., Lebauer, D. S., & Kooper, R. (2013). On improving the communication between models and data. *Plant, Cell and Environment*, *36*(9), 1575–1585. https://doi.org/10.1111/pce.12043

Dongarra, J., Getov, V., & Walsh, K. (2018). The 30th anniversary of the supercomputing conference: Bringing the future closer-supercomputing history and the immortality of now. *Computer*, *51*(10), 74–85. https://doi.org/10.1109/mc.2018.3971352

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., et al. (2017). ESA CCI soil moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment*, *203*, 185–215. https://doi.org/10.1016/j.rse.2017.07.001

Doucet, A., Godsill, S., & Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, *10*(3), 197–208. https://doi.org/10.1023/a:1008935410038

Dowd, M., & Joy, R. (2011). Estimating behavioral parameters in animal movement models using a state-augmented particle filter. *Ecology*, *92*(3), 568–575. https://doi.org/10.1890/10-0611.1

Draper, C. S., Reichle, R. H., De Lannoy, G. J. M., & Liu, Q. (2012). Assimilation of passive and active microwave soil moisture retrievals. *Geophysical Research Letters*, *39*. https://doi.org/10.1029/2011gl050655

Drévillon, M., Régnier, C., Lellouche, J.-M., Garric, G., Bricaud, C., Hernandez, O. (2018). *CMEMS-GLO-QUID-001-030, 1.2 edn., E.U. Copernicus marine service information [online]*. Retrieved from https://resources.marine.copernicus.eu/documents/QUID/CMEMS-GLO-QUID-001-030.pdf (Accessed: 28 December 2020).

Duncanson, L., Neuenschwander, A., Hancock, S., Thomas, N., Fatoyinbo, T., Simard, M., et al. (2020). Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California. *Remote Sensing of Environment*, *242*, 111779. https://doi.org/10.1016/j.rse.2020.111779

Dunne, S., & Entekhabi, D. (2005). An ensemble-based reanalysis approach to land data assimilation. *Water Resources Research*, *41*(2). https://doi.org/10.1029/2004wr003449

Durand, M., Molotch, N. P., & Margulis, S. A. (2008). A Bayesian approach to snow water equivalent reconstruction. *Journal of Geophysical Research*, *113*(D20). https://doi.org/10.1029/2008jd009894

Dutra, E., Balsamo, G., Viterbo, P., Miranda, P. M. A., Beljaars, A., Schar, C., & Elder, K. (2010). An improved snow scheme for the ECMWF land surface model: Description and offline validation. *Journal of Hydrometeorology*, *11*(4), 899–916. https://doi.org/10.1175/2010jhm1249.1

Eisen, K., & Plotkin, A. B. (2015). Forty years of forest measurements support steadily increasing aboveground biomass in a maturing, *Quercus*-dominant northeastern forest. *Journal of the Torrey Botanical Society*, *142*(2), 97–112. https://doi.org/10.3159/TORREY-D-14-00027.1

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology Evolution and Systematics*, *40*, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Emerick, A. A., & Reynolds, A. C. (2013). Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, *55*, 3–15. https://doi.org/10.1016/j.cageo.2012.03.011

Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, *53*(4), 343–367. https://doi.org/10.1007/s10236-003-0036-9

Evensen, G. (2018). Analysis of iterative ensemble smoothers for solving inverse problems. *Computational Geosciences*, *22*(3), 885–908. https://doi.org/10.1007/s10596-018-9731-y

Exbrayat, J.-F; Williams, M. (2018). *CARDAMOM panarctic retrievals 2000-2015, 2000-2015 (Dataset)*. National Centre for Earth Observation and School of GeoSciences. University of Edinburgh. https://doi.org/10.7488/ds/2334

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fei, S. L., Guo, Q. F., & Potter, K. (2016). Macrosystems ecology: Novel methods and new understanding of multi-scale patterns and processes. *Landscape Ecology*, *31*(1), 1–6. https://doi.org/10.1007/s10980-015-0315-0

Fennel, K., Gehlen, M., Brasseur, P., Brown, C. W., Ciavatta, S., Cossarini, G., et al. (2019). Advancing marine biogeochemical and ecosystem reanalyses and forecasts as tools for monitoring and managing ecosystem health. *Frontiers in Marine Science*, *6*, 89. https://doi.org/10.3389/fmars.2019.00089

Fer, I., Gardella, A. K., Shiklomanov, A. N., Serbin, S. P., De Kauwe, M. G., Raiho, A., et al. (2020). Beyond modeling: A roadmap to community cyberinfrastructure for ecological data-model integration. *Global Change Biology*, *27*, 13–26. https://doi.org/10.1111/gcb.15409

Firbank, L. G., Bertora, C., Blankman, D., Delle Vedove, G., Frenzel, M., Grignani, C., et al. (2017). Towards the co-ordination of terrestrial ecosystem protocols across European research infrastructures. *Ecology and Evolution*, *7*(11), 3967–3975. https://doi.org/10.1002/ece3.2997

Fontana, C., Brasseur, P., & Brankart, J. M. (2013). Toward a multivariate reanalysis of the North Atlantic Ocean biogeochemistry during 1998-2006 based on the assimilation of SeaWiFS chlorophyll data. *Ocean Science*, *9*(1), 37–56. https://doi.org/10.5194/os-9-37-2013

Ford, D., & Barciela, R. (2017). Global marine biogeochemical reanalyses assimilating two different sets of merged ocean colour products. *Remote Sensing of Environment*, *203*, 40–54. https://doi.org/10.1016/j.rse.2017.03.040

Forget, G., Campin, J. M., Heimbach, P., Hill, C. N., Ponte, R. M., & Wunsch, C. (2015). ECCO version 4: An integrated framework for non-linear inverse modeling and global ocean state estimation. *Geoscientific Model Development*, *8*(10), 3071–3104. https://doi.org/10.5194/gmd-8-3071-2015

Foster, D. R., Motzkin, G., & Slater, B. (1998). Land-use history as long-term broad-scale disturbance: regional forest dynamics in Central New England. *Ecosystems*, *1*(1), 96–119. https://doi.org/10.1007/s100219900008

Fox, A., Williams, M., Richardson, A. D., Cameron, D., Gove, J. H., Quaife, T., et al. (2009). The REFLEX project: Comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data. *Agricultural and Forest Meteorology*, *149*(10), 1597–1615. https://doi.org/10.1016/j.agrformet.2009.05.002

Fox, A. M., Hoar, T. J., Anderson, J. L., Arellano, A. F., Smith, W. K., Litvak, M. E., et al. (2018). Evaluation of a data assimilation system for land surface models using CLM4.5. *Journal of Advances in Modeling Earth Systems*, *10*(10), 2471–2494. https://doi.org/10.1029/2018ms001362

Friedlingstein, P, Meinshausen, M, Arora, V. K, Jones, C. D, Anav, A., Liddicoat, S. K., & Knutti, R. (2014). Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate*, *27*(2), 511–526. https://doi.org/10.1175/jcli-d-12-00579.1

Fuhrer, O., Chadha, T., Hoefler, T., Kwasniewski, G., Lapillonne, X., Leutwyler, D., et al. (2018). Near-global climate simulation at 1 km resolution: Establishing a performance baseline on 4888 GPUs with COSMO 5.0. *Geoscientific Model Development*, *11*(4), 1665–1681. https://doi.org/10.5194/gmd-11-1665-2018

Fujiwara, M., Wright, J. S., Manney, G. L., Gray, L. J., Anstey, J., Birner, T., et al. (2017). Introduction to the SPARC reanalysis intercomparison project (S-RIP) and overview of the reanalysis systems. *Atmospheric Chemistry and Physics*, *17*(2), 1417–1452. https://doi.org/10.5194/acp-17-1417-2017

Gámez-Virués, S., Perović, D. J., Gossner, M. M., Börschig, C., Blüthgen, N., de Jong, H., et al. (2015). Landscape simplification filters species traits and drives biotic homogenization. *Nature Communications*, *6*, 8568. https://doi.org/10.1038/ncomms9568

Gamon, J. A, Somers, B., Malenovsky, Z., Middleton, E. M., Rascher, U., & Schaepman, M. E (2019). Assessing vegetation function with imaging spectroscopy. *Surveys in Geophysics*, *40*(3), 489–513. https://doi.org/10.1007/s10712-019-09511-5

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gasper, F., Goergen, K., Shrestha, P., Sulis, M., Rihani, J., Geimer, M., & Kollet, S. (2014). Implementation and scaling of the fully coupled terrestrial systems modeling platform (TerrSysMP v1.0) in a massively parallel supercomputing environment - A case study on JUQUEEN (IBM Blue Gene/Q). *Geoscientific Model Development*, *7*(5), 2531–2543. https://doi.org/10.5194/gmd-7-2531-2014

GBIF. (2020). *What is GBIF? The Global Biodiversity Information Facility*. Retrieved from https://www.gbif.org/what-is-gbif

Gelaro, R., McCarty, W., Suarez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/jcli-d-16-0758.1

George, P. B. L., Lallias, D., Creer, S., Seaton, F. M., Kenny, J. G., Eccles, R. M., et al. (2019). Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. *Nature Communications*, *10*, 1107. https://doi.org/10.1038/s41467-019-09031-1

Gharamti, M. E., Tjiputra, J., Bethke, I., Samuelsen, A., Skjelvan, I., Bentsen, M., & Bertino, L. (2017). Ensemble data assimilation for ocean biogeochemical state and parameter estimation at different sites. *Ocean Modelling*, *112*, 65–89. https://doi.org/10.1016/j.ocemod.2017.02.006

Ghiggi, G., Humphrey, V., Seneviratne, S. I., & Gudmundsson, L. (2019). GRUN: An observation-based global gridded runoff dataset from 1902 to 2014. *Earth System Science Data*, *11*(4), 1655–1674. https://doi.org/10.5194/essd-11-1655-2019

Gholz, H. L., & Blood, E. R. (2016). Macro systems biology: Stimulating new perspectives on scaling in ecology. *Landscape Ecology*, *31*(1), 215–216. https://doi.org/10.1007/s10980-015-0300-7

Gibson, J., Kållberg, P. W., Uppala, S., Angeles, H., Nomura, A., & Serrano, E. (1997). *ERA Description*. Shinfield Park. Reading. Retrieved from https://www.ecmwf.int/node/9584

Giorgi, F., & Xue-Jie, G. (2018). Regional earth system modeling: Review and future directions. *Atmospheric and Oceanic Science Letters*, *11*(2), 189–197. https://doi.org/10.1080/16742834.2018.1452520

Girotto, M., De Lannoy, G. J. M., Reichle, R. H., & Rodell, M. (2016). Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model. *Water Resources Research*, *52*(5), 4164–4183. https://doi.org/10.1002/2015wr018417

Gomez-Hernandez, J. J., Sahuquillo, A., & Capilla, J. E. (1997). Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data - I. Theory. *Journal of Hydrology*, *203*(1–4), 162–174.

Green, J. L., Bohannan, B. J. M., & Whitaker, R. J. (2008). Microbial biogeography: From taxonomy to traits. *Science*, *320*(5879), 1039–1043. https://doi.org/10.1126/science.1153475

Gregg, W. W., & Rousseaux, C. S. (2019). Global ocean primary production trends in the modern ocean color satellite record (1998? 2015). *Environmental Research Letters*, *14*(12). https://doi.org/10.1088/1748-9326/ab4667

Gregory, R. D., van Strien, A., Vorisek, P., Meyling, A. W. G., Noble, D. G., Foppen, R. P. B., & Gibbons, D. W. (2005). Developing indicators for European birds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1454), 269–288. https://doi.org/10.1098/rstb.2004.1602

Grosse, M., Hierold, W., Ahlborn, M. C., & Helming, K. (2020). Long-term field experiments in Germany: Classification and spatial representation. *SOIL Discuss*, *6*, 1–32. https://doi.org/10.5194/soil-6-579-2020

Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., & Dorigo, W. (2019). Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology. *Earth System Science Data*, *11*(2), 717–739. https://doi.org/10.5194/essd-11-717-2019

Guerry, A. D., Polasky, S., Lubchenco, J., Chaplin-Kramer, R., Daily, G. C., Griffin, R., et al. (2015). Natural capital and ecosystem services informing decisions: From promise to practice. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(24), 7348–7355. https://doi.org/10.1073/pnas.1503751112

Haase, P., Tonkin, J. D., Stoll, S., Burkharde, B., Frenzel, M., Geijzendorffer, I. R., et al. (2018). The next generation of site-based long-term ecological monitoring: Linking essential biodiversity variables and ecosystem integrity. *The Science of the Total Environment*, *613*, 1376–1384. https://doi.org/10.1016/j.scitotenv.2017.08.111

Hanke, M., Redler, R., Holfeld, T., & Yastremsky, M. (2016). YAC 1.2.0: New aspects for coupling software in Earth system modelling. *Geoscientific Model Development*, *9*(8), 2755–2769. https://doi.org/10.5194/gmd-9-2755-2016

Hansen, M. C., Stehman, S. V., & Potapov, P. V. (2010). Quantification of global gross forest cover loss. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(19), 8650–8655. https://doi.org/10.1073/pnas.0912668107

Harada, Y., Kamahori, H., Kobayashi, C., Endo, H., Kobayashi, S., Ota, Y., et al. (2016). The JRA-55 reanalysis: Representation of atmospheric circulation and climate variability. *Journal of the Meteorological Society of Japan–Series II*, *94*(3), 269–302. https://doi.org/10.2151/jmsj.2016-015

Heffernan, J. B., Soranno, P. A., Angilletta, M. J., Buckley, L. B., Gruner, D. S., Keitt, T. H., et al. (2014). Macrosystems ecology: Understanding ecological patterns and processes at continental scales. *Frontiers in Ecology and the Environment*, *12*(1), 5–14. https://doi.org/10.1890/130017

Heimann, M., & Reichstein, M. (2008). Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature*, *451*(7176), 289–292. https://doi.org/10.1038/nature06591

Heinze, C., Eyring, V., Friedlingstein, P., Jones, C., Balkanski, Y., Collins, W., et al. (2019). ESD reviews: Climate feedbacks in the Earth system and prospects for their evaluation. *Earth System Dynamics*, *10*(3), 379–452. https://doi.org/10.5194/esd-10-379-2019

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Munoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hill, T. C., Ryan, E., & Williams, M. (2012). The use of $CO_2$ flux time series for parameter and carbon stock estimation in carbon cycle research. *Global Change Biology*, *18*(1), 179–193. https://doi.org/10.1111/j.1365-2486.2011.02511.x

Hobbs, W. R., Klekociuk, A. R., & Pan, Y. (2020). Validation of reanalysis Southern Ocean atmosphere trends using sea ice data. *Atmospheric Chemistry and Physics Discussions*, *20*, 1–17. https://doi.org/10.5194/acp-20-14757-2020

Hobeichi, S., Abramowitz, G., Evans, J., & Beck, H. E. (2019). Linear optimal runoff aggregate (LORA): A global gridded synthesis runoff product. *Hydrology and Earth System Sciences*, *23*(2), 851–870. https://doi.org/10.5194/hess-23-851-2019

Hobeichi, S., Abramowitz, G., Evans, J., & Ukkola, A. (2018). Derived optimal linear combination evapotranspiration (DOLCE): A global gridded synthesis ET estimate. *Hydrology and Earth System Sciences*, *22*(2), 1317–1336. https://doi.org/10.5194/hess-22-1317-2018

Hoffmann, L., Gunther, G., Li, D., Stein, O., Wu, X., Griessbach, S., et al. (2019). From ERA-Interim to ERA5: The considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. *Atmospheric Chemistry and Physics*, *19*(5), 3097–3124. https://doi.org/10.5194/acp-19-3097-2019

Hu, J., Fennel, K., Mattern, J. P., & Wilkin, J. (2012). Data assimilation with a local ensemble Kalman filter applied to a three-dimensional biological model of the middle Atlantic Bight. *Journal of Marine Systems*, *94*, 145–156. https://doi.org/10.1016/j.jmarsys.2011.11.016

Hudson, L. N., Newbold, T., Contu, S., Hill, S. L. L., Lysenko, I., De Palma, A., et al. (2017). The database of the PREDICTS (projecting responses of ecological diversity in changing terrestrial systems) project. *Ecology and Evolution*, *7*(1), 145–188.

Hulley, G., Hook, S., Fisher, J., & Lee, C. (2017). Ecostress, a Nasa Earth-ventures instrument for studying links between the water cycle and plant health over the diurnal cycle. In *IEEE International Geoscience and Remote Sensing Symposium (Igarss)* (pp. 5494–5496). https://doi.org/10.1109/igarss.2017.8128248

Humphrey, V., & Gudmundsson, L. (2019). GRACE-REC: A reconstruction of climate-driven water storage changes over the last century. *Earth System Science Data*, *11*(3), 1153–1170. https://doi.org/10.5194/essd-11-1153-2019

Huntzinger, D. N., Michalak, A. M., Schwalm, C., Ciais, P., King, A. W., Fang, Y., et al. (2017). Uncertainty in the response of terrestrial carbon sink to environmental drivers undermines carbon-climate feedback predictions. *Scientific Reports*, *7*. https://doi.org/10.1038/s41598-017-03818-2

Ishizaka, J. (1990). Coupling of coastal zone color scanner data to a physical-biological model of the southeastern U.S. continental shelf ecosystem: 2. An Eulerian model. *Journal of Geophysical Research*, *95*(C11), 20183–20199. https://doi.org/10.1029/JC095iC11p20183

Jetz, W., McGeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M., et al. (2019). Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology & Evolution*, *3*(4), 539–551. https://doi.org/10.1038/s41559-019-0826-1

Ji, M., Kumar, A., & Leetmaa, A. (1994). A multiseason climate forecast system at the national-meteorological-center. *Bulletin of the American Meteorological Society*, *75*(4), 569–577. https://doi.org/10.1175/1520-0477(1994)075<0569:amcfsa>2.0.co;2

Jimenez, C., Martens, B., Miralles, D. M., Fisher, J. B., Beck, H. E., & Fernandez-Prieto, D. (2018). Exploring the merging of the global land evaporation WACMOS-ET products based on local tower measurements. *Hydrology and Earth System Sciences*, *22*(8), 4513–4533. https://doi.org/10.5194/hess-22-4513-2018

Jockel, P., Kerkweg, A., Pozzer, A., Sander, R., Tost, H., Riede, H., et al. (2010). Development cycle 2 of the modular Earth submodel system (MESSy2). *Geoscientific Model Development*, *3*(2), 717–752. https://doi.org/10.5194/gmd-3-717-2010

Jones, E. M., Baird, M. E., Mongin, M., Parslow, J., Skerratt, J., Lovell, J., et al. (2016). Use of remote-sensing reflectance to constrain a data assimilating marine biogeochemical model of the Great Barrier Reef. *Biogeosciences*, *13*(23), 6441–6469. https://doi.org/10.5194/bg-13-6441-2016

Jones, L. A., Kimball, J. S., Reichle, R. H., Madani, N., Glassy, J., Ardizzone, J. V., et al. (2017). The SMAP level 4 carbon product for monitoring ecosystem land-atmosphere $CO_2$ exchange. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(11), 6517–6532. https://doi.org/10.1109/tgrs.2017.2729343

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*(1), 35–45. https://doi.org/10.1115/1.3662552

Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–471. https://doi.org/10.1175/1520-0477(1996)077<0437:tnyrp>2.0.co;2

Kaminski, T., Knorr, W., Scholze, M., Gobron, N., Pinty, B., Giering, R., & Mathieu, P. P. (2012). Consistent assimilation of MERIS FAPAR and atmospheric $CO_2$ into a terrestrial vegetation model and interactive mission benefit analysis. *Biogeosciences*, *9*(8), 3173–3184. https://doi.org/10.5194/bg-9-3173-2012

Kamp, J., Reinhard, A., Frenzel, M., Kampfer, S., Trappe, J., & Holzel, N. (2018). Farmland bird responses to land abandonment in Western Siberia. *Agriculture, Ecosystems & Environment*, *268*, 61–69. https://doi.org/10.1016/j.agee.2018.09.009

Kattge, J., Bonisch, G., Diaz, S., Lavorel, S., Prentice, I. C., Leadley, P., et al. (2020). TRY plant trait database - Enhanced coverage and open access. *Global Change Biology*, *26*(1), 119–188.

Kattge, J., Diaz, S., Lavorel, S., Prentice, C., Leadley, P., Bonisch, G., et al. (2011). TRY - A global database of plant traits. *Global Change Biology*, *17*(9), 2905–2935.

Kaufman, D. E., Friedrichs, M. A. M., Hemmings, J. C. P., & Smith, W. O. (2018). Assimilating bio-optical glider data during a phytoplankton bloom in the southern Ross Sea. *Biogeosciences*, *15*(1), 73–90. https://doi.org/10.5194/bg-15-73-2018

Kearney, M., & Porter, W. (2009). Mechanistic niche modelling: Combining physiological and spatial data to predict species' ranges. *Ecology Letters*, *12*(4), 334–350. https://doi.org/10.1111/j.1461-0248.2008.01277.x

Keller, M., Schimel, D. S., Hargrove, W. W., & Hoffman, F. M. (2008). A continental strategy for the national ecological observatory network. *Frontiers in Ecology and the Environment*, *6*(5), 282–284. https://doi.org/10.1890/1540-9295(2008)6[282:acsftn]2.0.CO;2

Kissling, W. D., Walls, R., Bowser, A., Jones, M. O., Kattge, J., Agosti, D., et al. (2018). Towards global data products of Essential Biodiversity Variables on species traits. *Nature Ecology & Evolution*, *2*(10), 1531–1540. https://doi.org/10.1038/s41559-018-0667-3

Kitchin, R., & McArdle, G. (2016). What makes big data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1), 1–10. https://doi.org/10.1177/2053951716631130

Kleist, D. T., Parrish, D. F., Derber, J. C., Treadon, R., Errico, R. M., & Yang, R. H. (2009). Improving incremental balance in the GSI 3DVAR analysis system. *Monthly Weather Review*, *137*(3), 1046–1060. https://doi.org/10.1175/2008mwr2623.1

Knape, J., & de Valpine, P. (2012). Fitting complex population models by combining particle filters with Markov chain Monte Carlo. *Ecology*, *93*(2), 256–263. https://doi.org/10.1890/11-0797.1

Knorr, W. (2000). Annual and interannual $CO_2$ exchanges of the terrestrial biosphere: Process-based simulations and uncertainties. *Global Ecology and Biogeography*, *9*(3), 225–252. https://doi.org/10.1046/j.1365-2699.2000.00159.x

Knorr, W., & Heimann, M. (1995). Impact of drought stress and other factors on seasonal land biosphere $Co_2$ exchange studied through an atmospheric tracer transport model. *Tellus Series B Chemical and Physical Meteorology*, *47*(4), 471–489. https://doi.org/10.1034/j.1600-0889.47.issue4.7.x

Knorr, W., & Kattge, J. (2005). Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. *Global Change Biology*, *11*(8), 1333–1351. https://doi.org/10.1111/j.1365-2486.2005.00977.x

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan–Series II*, *93*(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Koffi, E. N., Rayner, P. J., Norton, A. J., Frankenberg, C., & Scholze, M. (2015). Investigating the usefulness of satellite-derived fluorescence data in inferring gross primary productivity within the carbon cycle data assimilation system. *Biogeosciences*, *12*(13), 4067–4084. https://doi.org/10.5194/bg-12-4067-2015

Kollet, S. J., & Maxwell, R. M. (2008). Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model. *Water Resources Research*, *44*(2). https://doi.org/10.1029/2007wr006004

Kossin, J. P. (2015). Validating atmospheric reanalysis data using tropical cyclones as thermometers. *Bulletin of the American Meteorological Society*, *96*(7), 1089–1096. https://doi.org/10.1175/bams-d-14-00180.1

Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., & Kumar, P. (2000). A catchment-based approach to modeling land surface processes in a general circulation model 1. Model structure. *Journal of Geophysical Research-Atmospheres*, *105*(D20), 24809–24822. https://doi.org/10.1029/2000jd900327

Krajsek, K., Comito, C., Götz, M., Hagemeier, B., Knechtges, P., & Siggel, M. (2018). The Helmholtz Analytics Toolkit (HeAT) - A scientific big data library for HPC. Extreme data – demands, technologies, and services, Jülich, 18.-19. *IAS Series*, *40*, 57–60. Retrieved from http://hdl.handle.net/2128/22029

Kubiszewski, I., Costanza, R., Anderson, S., & Sutton, P. (2017). The future value of ecosystem services: Global scenarios and national implications. *Ecosystem Services*, *26*, 289–301. https://doi.org/10.1016/j.ecoser.2017.05.004

Kumar, S. V., Dong, J. R., Peters-Lidard, C. D., Mocko, D., & Gomez, B. (2017). Role of forcing uncertainty and background model error characterization in snow data assimilation. *Hydrology and Earth System Sciences*, *21*(6), 2637–2647. https://doi.org/10.5194/hess-21-2637-2017

Kumar, S. V., Mocko, D. M., Wang, S. G., Peters-Lidard, C. D., & Borak, J. (2019). Assimilation of remotely sensed leaf area index into the Noah-MP land surface model: Impacts on water and carbon fluxes and states over the continental United States. *Journal of Hydrometeorology*, *20*(7), 1359–1377. https://doi.org/10.1175/jhm-d-18-0237.1

Kumar, S. V., Zaitchik, B. F., Peters-Lidard, C. D., Rodell, M., Reichle, R., Li, B. L., et al. (2016). Assimilation of gridded GRACE terrestrial water storage estimates in the North American land data assimilation system. *Journal of Hydrometeorology*, *17*(7), 1951–1972. https://doi.org/10.1175/jhm-d-15-0157.1

Kurtz, W., He, G. W., Kollet, S. J., Maxwell, R. M., Vereecken, H., & Franssen, H. J. H. (2016). TerrSysMP-PDAF version 1.0: A modular high-performance data assimilation framework for an integrated land surface-subsurface model. *Geoscientific Model Development*, *9*(4), 1341–1360. https://doi.org/10.5194/gmd-9-1341-2016

Lal, R. (2015). Restoring soil quality to mitigate soil degradation. *Sustainability*, *7*(5), 5875–5895. https://doi.org/10.3390/su7055875

Laloyaux, P., de Boisseson, E., Balmaseda, M., Bidlot, J. R., Broennimann, S., Buizza, R., et al. (2018). CERA-20C: A coupled reanalysis of the twentieth century. *Journal of Advances in Modeling Earth Systems*, *10*(5), 1172–1195. https://doi.org/10.1029/2018ms001273

Lausch, A., Baade, J., Bannehr, L., Borg, E., Bumberger, J., Chabrilliat, S., et al. (2019). Linking remote sensing and geodiversity and their traits relevant to biodiversity-part I: Soil characteristics. *Remote Sensing*, *11*(20). https://doi.org/10.3390/rs11202356

Lausch, A., Bannehr, L., Beckmann, M., Boehm, C., Feilhauer, H., Hacker, J. M., et al. (2016). Linking Earth observation and taxonomic, structural and functional biodiversity: Local to ecosystem perspectives. *Ecological Indicators*, *70*, 317–339. https://doi.org/10.1016/j.ecolind.2016.06.022

Lausch, A., Borg, E., Bumberger, J., Dietrich, P., Heurich, M., Huth, A., et al. (2018). Understanding forest health with remote sensing, Part III: Requirements for a scalable multi-source forest health monitoring network based on data science approaches. *Remote Sensing*, *10*(7). https://doi.org/10.3390/rs10071120

Lavric, J. V., Juurola, E., Vermeulen, A. T., & Kutsch, W. L. (2016). *Integrated carbon observation system (ICOS) - a domain-overarching long-term research infrastructure for the future*. Retrieved from https://ui.adsabs.harvard.edu/abs/2016AGUFMGC21C1117L

Lawrence, B. N., Rezny, M., Budich, R., Bauer, P., Behrens, J., Carter, M., et al. (2018). Crossing the chasm: How to develop weather and climate models for next generation computers? *Geoscientific Model Development*, *11*(5), 1799–1821. https://doi.org/10.5194/gmd-11-1799-2018

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., et al. (2019). The community land model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4245–4287

Lebauer, D. S., Wang, D, Richter, K. T., Davidson, C. C., & Dietze, M. C. (2013). Facilitating feedbacks between field measurements and ecosystem models. *Ecological Monographs*, *83*(2), 133–154. https://doi.org/10.1890/12-0137.1

Lees, K. J., Quaife, T., Artz, R. R. E., Khomik, M., & Clark, J. M. (2018). Potential for using remote sensing to estimate carbon fluxes across northern peatlands - A review. *The Science of the Total Environment*, *615*, 857–874. https://doi.org/10.1016/j.scitotenv.2017.09.103

Lellouche, J.-M., Le Galloudec, O., Greiner, E., Garric, G., Regnier, C., Drevillon, M., et al. (2018). *The Copernicus Marine Environment Monitoring Service global ocean 1/12° physical reanalysis GLORYS12V1: description and quality assessment. In 20th EGU General Assembly, EGU2018, Proceedings from the conference* (Vol. 2018).

Leutwyler, D., Fuhrer, O., Lapillonne, X., Luthi, D., & Schar, C. (2016). Towards European-scale convection-resolving climate simulations with GPUs: A study with COSMO 4.19. *Geoscientific Model Development*, *9*(9), 3393–3412. https://doi.org/10.5194/gmd-9-3393-2016

Li, B. L., Rodell, M., Kumar, S., Beaudoing, H. K., Getirana, A., Zaitchik, B. F., et al. (2019). Global GRACE data assimilation for groundwater and drought monitoring: Advances and challenges. *Water Resources Research*, *55*(9), 7564–7586. https://doi.org/10.1029/2018wr024618

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land-surface water and energy fluxes for general-circulation models. *Journal of Geophysical Research*, *99*(D7), 14415–14428. https://doi.org/10.1029/94jd00483

Lindstrom, E., Gunn, J., Fischer, A., McCurdy, A., & Glover, L. K., & Members of the Task Team. (2010). *A framework for ocean observing*. UNESCO.

Lorenc, A. C., Bowler, N. E., Clayton, A. M., Pring, S. R., & Fairbairn, D. (2015). Comparison of hybrid-4DEnVar and Hybrid-4DVar data assimilation methods for global NWP. *Monthly Weather Review*, *143*(1), 212–229. https://doi.org/10.1175/mwr-d-14-00195.1

Lorenz, C., & Kunstmann, H. (2012). The hydrological cycle in three state-of-the-art reanalyses: Intercomparison and performance analysis. *Journal of Hydrometeorology*, *13*(5), 1397–1420. https://doi.org/10.1175/jhm-d-11-088.1

Lorimer, C. G. (1984). Development of the red maple understory in northeastern oak forests. *Forest Science*, *30*(1), 3–22. https://doi.org/10.1093/forestscience/30.1.3

Makiola, A., Compson, Z. G., Baird, D. J., Barnes, M. A., Boerlijst, S. P., Bouchez, A., et al. (2020). Key questions for next-generation biomonitoring. *Frontiers in Environmental Science*, *7*(197). https://doi.org/10.3389/fenvs.2019.00197

Martin-Fernandez, L., Gilioli, G., Lanzarone, E., Miguez, J., Pasquali, S., Ruggeri, F., et al. (2014). A Rao-Blackwellized particle filter for joint parameter estimation and biomass tracking in a stochastic predator-prey system. *Mathematical Biosciences and Engineering*, *11*(3), 573–597. https://doi.org/10.3934/mbe.2014.11.573

Matheny, A. M., Mirfenderesgi, G., & Bohrer, G. (2017). Trait-based representation of hydrological functional properties of plants in weather and ecosystem models. *Plant Diversity*, *39*(1), 1–12. https://doi.org/10.1016/j.pld.2016.10.001

Maxwell, R. M., & Kollet, S. J. (2008). Interdependence of groundwater dynamics and land-energy feedbacks under climate change. *Nature Geoscience*, *1*(10), 665–669. https://doi.org/10.1038/ngeo315

Medvigy, D., Wofsy, S. C., Munger, J. W., Hollinger, D. Y., & Moorcroft, P. R. (2009). Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem demography model version 2. *Journal of Geophysical Research*, *114*(G1). https://doi.org/10.1029/2008jg000812

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092. https://doi.org/10.1063/1.1699114

Miguel, M., Mathias, D., David, F., Martin, H., Mirco, M., Jens-Christian, S., et al. (2017). *Essential ecosystem variables*.

Mirtl, M., Borer, T. E., Djukic, I., Forsius, M., Haubold, H., Hugo, W., et al. (2018). Genesis, goals and achievements of long-term ecological research at the global scale: A critical review of ILTER and future directions. *The Science of the Total Environment*, *626*, 1439–1462. Retrieved from http://www.sciencedirect.com/science/article/pii/S0048969717334204

Montzka, C., Pauwels, V. R. N., Franssen, H. J. H., Han, X. J., & Vereecken, H. (2012). Multivariate and multiscale data assimilation in terrestrial systems: A Review. *Sensors*, *12*(12), 16291–16333.

Moreno-Martinez, A., Camps-Valls, G., Kattge, J., Robinson, N., Reichstein, M., van Bodegom, P., et al. (2018). A methodology to derive global maps of leaf traits using remote sensing and climate data. *Remote Sensing of Environment*, *218*, 69–88.

Mosegaard, K., & Sambridge, M. (2002). Monte Carlo analysis of inverse problems. *Inverse Problems*, *18*(3), R29–R54.

Mosegaard, K., & Tarantola, A. (1995). Monte-Carlo Sampling of Solutions to Inverse Problems. *Journal of Geophysical Research*, *100*(B7), 12431–12447.

Motesharrei, S., Rivas, J., Kalnay, E., Asrar, G. R., Busalacchi, A. J., Cahalan, R. F., et al. (2016). Modeling sustainability: Population, inequality, consumption, and bidirectional coupling of the Earth and human systems. *National Science Review*, *3*(4), 470–494.

Muller, F., Hoffmann-Kroll, R., & Wiggering, H. (2000). Indicating ecosystem integrity - Theoretical concepts and environmental requirements. *Ecological Modelling*, *130*(1–3), 13–23.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., et al. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications, *Earth System Science Data Discussions* [preprint], https://doi.org/10.5194/essd-2021-82

Naz, B. S., Kollet, S., Franssen, H. J. H., Montzka, C., & Kurtz, W. (2020). A 3 km spatially and temporally consistent European daily soil moisture reanalysis from 2000 to 2015. *Scientific Data*, *7*(1).

Ndehedehe, C. E., Awange, J. L., Agutu, N. O., & Okwuashi, O. (2018). Changes in hydro-meteorological conditions over tropical West Africa (1980-2015) and links to global climate. *Global and Planetary Change*, *162*, 321–341.

Nerger, L., Danilov, S., Hiller, W., & Schroter, J. (2006). Using sea-level data to constrain a finite-element primitive-equation ocean model with a local SEIK filter. *Ocean Dynamics*, *56*(5–6), 634–649.

Nerger, L., & Gregg, W. W. (2008). Improving assimilation of SeaWiFS data by the application of bias correction with a local SEIK filter. *Journal of Marine Systems*, *73*(1–2), 87–102.

Nerger, L., & Hiller, W. (2013). Software for ensemble-based data assimilation systems-Implementation strategies and scalability. *Computers & Geosciences*, *55*, 110–118.

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, *116*(D12).

Noilhan, J., & Mahfouf, J. F. (1996). The ISBA land surface parameterisation scheme. *Global and Planetary Change*, *13*(1–4), 145–159.

Oddershede, A., Violle, C., Baattrup-Pedersen, A., Svenning, J. C., & Damgaard, C. (2019). Early dynamics in plant community trait responses to a novel, more extreme hydrological gradient. *Journal of Plant Ecology*, *12*(2), 327–335.

Oleson, K. W., Niu, G.-Y., Yang, Z. L., Lawrence, D. M., Thornton, P. E., Lawrence, P. J., et al., (2007), *Technical description of the community land model (CLM 3.5)*, NCAR.

Oleson, K. W., Niu, G. Y., Yang, Z. L., Lawrence, D. M., Thornton, P. E., Lawrence, P. J., et al. (2008). Improvements to the Community Land Model and their impact on the hydrological cycle. *Journal of Geophysical Research*, *113*(G1).

Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., et al. (2007). The JRA-25 reanalysis. *Journal of the Meteorological Society of Japan–Series II*, *85*(3), 369–432.

Overpeck, J. T., Meehl, G. A., Bony, S., & Easterling, D. R. (2011). Climate data challenges in the 21st century. *Science*, *331*(6018), 700–702.

Palmer, T. N. (2016). A personal perspective on modelling the climate system. *Proceedings of the Royal Society a-Mathematical Physical and Engineering Sciences*, *472*(2188).

Pan, M., Sahoo, A. K., Troy, T. J., Vinukollu, R. K., Sheffield, J., & Wood, E. F. (2012). Multisource estimation of long-term terrestrial water budget for major global river basins. *Journal of Climate*, *25*(9), 3191–3206.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., et al. (2006). Towards a standardized processing of net ecosystem exchange measured with eddy covariance technique: Algorithms and uncertainty estimation. *Biogeosciences*, *3*(4), 571–583. Retrieved from https://bg.copernicus.org/articles/3/571/2006/

Penny, S. G., Akella, S., Balmaseda, M. A., Browne, P., Carton, J. A., Chevallier, M., et al. (2019). Observational needs for improving ocean and coupled reanalysis, S2S prediction, and decadal prediction. *Frontiers in Marine Science*, *6*.

Penny, S. G., & Miyoshi, T. (2016). A local particle filter for high-dimensional geophysical systems. *Nonlinear Processes in Geophysics*, *23*(5), 391–405.

Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., et al. (2013). Essential Biodiversity Variables. *Science*, *339*(6117), 277–278.

Peylin, P., Bacour, C., MacBean, N., Leonard, S., Rayner, P., Kuppel, S., et al. (2016). A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle. *Geoscientific Model Development*, *9*(9), 3321–3346.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259.

Planning for First Garp Global Experiment (Fgge). (1972). *Bulletin of the American Meteorological Society*, 53(12), pp. 1163–1164).

Poli, P., Hans, H., Berrisford, P., & other, a. (2015). *ERA-20C deterministic*, 126(4). Retrieved from Shinfield Park. Reading. Retrieved from https://www.ecmwf.int/node/11700

Post, W. M., & Pastor, J. (1996). Linkages—an individual-based forest ecosystem model. *Climatic Change*, *34*(2), 253–261. https://doi.org/10.1007/BF00224636

Poterjoy, J. (2016). A localized particle filter for high-dimensional nonlinear systems. *Monthly Weather Review*, *144*(1), 59–76.

Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., et al. (2017). The weather research and forecasting model: Overview, system efforts, and future directions. *Bulletin of the American Meteorological Society*, *98*(8), 1717–1737.

Pradhan, H. K., Volker, C., Losa, S. N., Bracher, A., & Nerger, L. (2019). Assimilation of global total chlorophyll OC-CCI data and its impact on individual phytoplankton fields. *Journal of Geophysical Research: Oceans*, *124*(1), 470–490.

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., et al. (2015). A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of Geophysics*, *53*(2), 323–361.

Proenca, V., Martin, L. J., Pereira, H. M., Fernandez, M., McRae, L., Belnap, J., et al. (2017). Global biodiversity monitoring: From data sources to essential biodiversity variables. *Biological Conservation*, *213*, 256–263.

Pulido, M., & van Leeuwen, P. J. (2019). Sequential Monte Carlo with kernel embedded mappings: Themapping particle filter. *Journal of Computational Physics*, *396*, 400–415.

Quegan, S., Toan, L. T., Chave, J., Dall, J., Exbrayat, J. F., Minh, D. H. T., et al. (2019). The European space agency BIOMASS mission: Measuring forest above-ground biomass from space. *Remote Sensing of Environment*, *227*, 44–60.

Raddatz, T. J., Reick, C. H., Knorr, W., Kattge, J., Roeckner, E., Schnur, R., et al. (2007). Will the tropical land biosphere dominate the climate-carbon cycle feedback during the twenty-first century? *Climate Dynamics*, *29*(6), 565–574.

Raiho, A., Dietze, M., Dawson, A., Rollinson, C. R., Tipton, J., & McLachlan, J. (2020). *Determinants of Predictability in Multi-decadal Forest Community and Carbon Dynamics* (Vol. 2005). Retrieved from https://www.biorxiv.org/content/biorxiv/early/2020/05/07/2020.05.05.079871.full.pdf

Rakhimberdiev, E., van den Hout, P. J., Brugge, M., Spaans, B., & Piersma, T. (2015). Seasonal mortality and sequential density dependence in a migratory bird. *Journal of Avian Biology*, *46*(4), 332–341.

Regos, A., Gagne, L., Alcaraz-Segura, D., Honrado, J. P., & Dominguez, J. (2019). Effects of species traits and environmental predictors on performance and transferability of ecological niche models. *Scientific Reports*, *9*.

Reichle, R., De Lannoy, G., Liu, Q., Ardizzone, J., Kimball, J., & Koster, R. (2016). Smap Level 4 Surface and Root Zone Soil Moisture. In *2016 IEEE international Geoscience and Remote sensing Symposium (Igarss)*, (pp. 136–138).

Reichle, R. H., De Lannoy, G. J. M., Forman, B. A., Draper, C. S., & Liu, Q. (2014). Connecting satellite observations with water cycle variables through land data assimilation: Examples using the NASA GEOS-5 LDAS. *Surveys in Geophysics*, *35*(3), 577–606.

Reichle, R. H., De Lannoy, G. J. M., Liu, Q., Ardizzone, J. V., Colliander, A., Conaty, A., et al. (2017a). Assessment of the SMAP Level-4 surface and root-zone soil moisture product using in situ measurements. *Journal of Hydrometeorology*, *18*(10), 2621–2645.

Reichle, R. H., Draper, C. S., Liu, Q., Girotto, M., Mahanama, S. P. P., Koster, R. D., & De Lannoy, G. J. M. (2017b). Assessment of MERRA-2 land surface hydrology estimates. *Journal of Climate*, *30*(8), 2937–2960.

Reichle, R. H., Koster, R. D., De Lannoy, G. J. M., Forman, B. A., Liu, Q., Mahanama, S. P. P., & Toure, A. (2011). Assessment and enhancement of merra land surface hydrology estimates. *Journal of Climate*, *24*(24), 6322–6338.

Reichle, R. H., Liu, Q., Koster, R. D., Crow, W., De Lannoy, G. J. M., Kimball, J. S., et al. (2019). Version 4 of the SMAP level-4 soil moisture algorithm and data product. *Journal of Advances in Modeling Earth Systems*, *11*(10), 3106–3130.

Reichle, R. H., McLaughlin, D. B., & Entekhabi, D. (2002). Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, *130*(1), 103–114.

Reichstein, M., Papale, D., Valentini, R., Aubinet, M., Bernhofer, C., Knohl, A., et al. (2007). Determinants of terrestrial ecosystem carbon balance inferred from European eddy covariance flux sites. *Geophysical Research Letters*, *34*(1). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GL027880

Reick, C. H., Raddatz, T., Brovkin, V., & Gayler, V. (2013). Representation of natural and anthropogenic land cover change in MPI-ESM. *Journal of Advances in Modeling Earth Systems*, *5*(3), 459–482.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., et al. (2011). MERRA: NASA's Modern-Era retrospective analysis for research and applications. *Journal of Climate*, *24*(14), 3624–3648.

Rodell, M., Beaudoing, H. K., L'Ecuyer, T. S., Olson, W. S., Famiglietti, J. S., Houser, P. R., et al. (2015). The observed state of the water cycle in the early twenty-first century. *Journal of Climate*, *28*(21), 8289–8318.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., et al. (2004). The global land data assimilation system. *Bulletin of the American Meteorological Society*, *85*(3), 381

Rosa, I. M. D., Purvis, A., Alkemade, R., Chaplin-Kramer, R., Ferrier, S., Guerra, C. A., et al. (2020). Challenges in producing policy-relevant global scenarios of biodiversity and ecosystem services. *Global Ecology and Conservation*, *22*, e00886. Retrieved from http://www.sciencedirect.com/science/article/pii/S2351989419306183

Rotzer, K., Montzka, C., & Vereecken, H. (2015). Spatio-temporal variability of global soil moisture products. *Journal of Hydrology*, *522*, 187–202.

Ryan, C. M., Hill, T., Woollen, E., Ghee, C., Mitchard, E., Cassells, G., et al. (2012). Quantifying small-scale deforestation and forest degradation in African woodlands using radar imagery. *Global Change Biology*, *18*(1), 243–257.

Saha, S., Moorthi, S., Pan, H. L., Wu, X. R., Wang, J. D., Nadiga, S., et al. (2010). The Ncep climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, *91*(8), 1015–1057.

Sahoo, A. K., Pan, M., Troy, T. J., Vinukollu, R. K., Sheffield, J., & Wood, E. F. (2011). Reconciling the global terrestrial water budget using satellite remote sensing. *Remote Sensing of Environment*, *115*(8), 1850–1865.

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*.

Schartau, M., Landry, M. R., & Armstrong, R. A. (2010). Density estimation of plankton size spectra: A reanalysis of IronEx II data. *Journal of Plankton Research*, *32*(8), 1167–1184.

Schiffers, K., Schurr, F. M., Travis, J. M. J., Duputie, A., Eckhart, V. M., Lavergne, S., et al. (2014). Landscape structure and genetic architecture jointly impact rates of niche evolution. *Ecography*, *37*(12), 1218–1229.

Schimel, D., Pavlick, R., Fisher, J. B., Asner, G. P., Saatchi, S., Townsend, P., et al. (2015). Observing terrestrial ecosystems and the carbon cycle from space. *Global Change Biology*, *21*(5), 1762–1776.

Schimel, D., Schneider, F. D., Bloom, A., Bowman, K., Cawse-Nicholson, K., Elder, C., et al. (2019). Flux towers in the sky: Global ecology from space. *New Phytologist*, *224*(2), 570–584.

Schnase, J. L., Lee, T. J., Mattmann, C. A., Lynnes, C. S., Cinquini, L., Ramirez, P. M., et al. (2016). Big data challenges in climate science improving the next-generation cyberinfrastructure. *IEEE Geoscience and Remote Sensing Magazine*, *4*(3), 10–22.

Schneider, F. D., Morsdorf, F., Schmid, B., Petchey, O. L., Hueni, A., Schimel, D. S., & Schaepman, M. E. (2017). Mapping functional diversity from remotely sensed morphological and physiological forest traits. *Nature Communications*, *8*. 1441.

Scholes, R. J., Mace, G. M., Turner, W., Geller, G. N., Jurgens, N., Larigauderie, A., et al. (2008). Ecology - Toward a global biodiversity observing system. *Science*, *321*(5892), 1044–1045.

Scholze, M., Buchwitz, M., Dorigo, W., Guanter, L., & Shaun, Q. G. (2017). Reviews and syntheses: Systematic Earth observations for use in terrestrial carbon cycle data assimilation systems. *Biogeosciences*, *14*(14), 3401–3429.

Schulthess, T. C., Bauer, P., Wedi, N., Fuhrer, O., Hoefler, T., & Schar, C. (2019). Reflecting on the goal and baseline for exascale computing: A roadmap based on weather and climate simulations. *Computing in Science & Engineering*, *21*(1), 30–41.

Shrestha, P., Sulis, M., Masbou, M., Kollet, S., & Simmer, C. (2014a). A Scale-consistent terrestrial systems modeling platform based on COSMO, CLM, and ParFlow. *Monthly Weather Review*, *142*(9), 3466–3483.

Shrestha, R. R., Schnorbus, M. A., Werner, A. T., & Zwiers, F. W. (2014b). Evaluating hydroclimatic change signals from statistically and dynamically downscaled GCMs and hydrologic models. *Journal of Hydrometeorology*, *15*(2), 844–860.

Shugart, H. H., Jr., & West, D. C. (1980). Forest succession models. *BioScience*, *30*(5), 308–313. https://doi.org/10.2307/1307854

Shugart, H. H., & Smith, T. M. (1996). A review of forest patch models and their application to global change research. *Climatic Change*, *34*(2), 131–153. https://doi.org/10.1007/BF00224626

Simon, E., Samuelsen, A., Bertino, L., & Mouysset, S. (2015). Experiences in multiyear combined state-parameter estimation with an ecosystem model of the North Atlantic and Arctic Oceans using the Ensemble Kalman Filter. *Journal of Marine Systems*, *152*, 1–17.

Sitzia, T., Campagnaro, T., Kotze, D. J., Nardi, S., & Ertani, A. (2018). The invasion of abandoned fields by a major alien tree filters understory plant traits in novel forest ecosystems. *Scientific Reports*, *8*, 8410.

Skakala, J., Bruggeman, J., Brewin, R. J. W., Ford, D. A., & Ciavatta, S. (2020). Improved Representation of Underwater Light Field and Its Impact on Ecosystem Dynamics: A Study in the North Sea. *Journal of Geophysical Research: Oceans*, *125*(7).

Skakala, J., Ford, D., Brewin, R. J. W., McEwan, R., Kay, S., Taylor, B., et al. (2018). The assimilation of phytoplankton functional types for operational forecasting in the northwest European shelf. *Journal of Geophysical Research: Oceans*, *123*(8), 5230–5247.

Skákala, J., Ford, D., Bruggeman, J., Hull, T., Kaiser, J., King, R. R., et al. (2021). Towards a multi-platform assimilative system for North Sea biogeochemistry. *Journal of Geophysical Research: Oceans*, *126*(4), e2020JC016649. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JC016649

Slivinski, L. C., Compo, G. P., Sardeshmukh, P. D., Whitaker, J. S., McColl, C., Allan, R. J., et al. (2021). An evaluation of the performance of the twentieth century reanalysis version 3. *Journal of Climate*, *34*(4), 1417–1438. https://doi.org/10.1175/jcli-d-20-0505.1

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., et al. (2019). Towards a more reliable historical reanalysis: Improvements for version 3 of the twentieth century reanalysis system. *Quarterly Journal of the Royal Meteorological Society*, *145*(724), 2876–2908.

Smith, B., Prentice, I. C., & Sykes, M. T. (2001). Representation of vegetation dynamics in the modelling of terrestrial ecosystems: Comparing two contrasting approaches within European climate space. *Global Ecology and Biogeography*, *10*(6), 621–637. https://doi.org/10.1046/j.1466-822x.2001.t01-1-00256.x

Song, H., Edwards, C. A., Moore, A. M., & Fiechter, J. (2016). Data assimilation in a coupled physical-biogeochemical model of the California current system using an incremental lognormal 4-dimensional variational approach: Part 2-Joint physical and biological data assimilation twin experiments. *Ocean Modelling*, *106*, 146–158.

Staniczenko, P. P. A., Sivasubramaniam, P., Suttle, K. B., & Pearson, R. G. (2017). Linking macroecology and community ecology: Refining predictions of species distributions using biotic interaction networks. *Ecology Letters*, *20*(6), 693–707.

Stavros, E. N., Schimel, D., Pavlick, R., Serbin, S., Swann, A., Duncanson, L., et al. (2017). ISS observations offer insights into plant function. *Nature Ecology & Evolution*, *1*(7), 194.

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., et al. (2019). DYAMOND: The DYnamics of the Atmospheric general circulation modeled on non-hydrostatic domains. *Progress in Earth and Planetary Science*, *6*(1).

Storto, A., Yang, C. X., & Masina, S. (2017). Constraining the global ocean heat content through assimilation of CERES-derived TOA energy imbalance estimates. *Geophysical Research Letters*, *44*(20), 10520–10529.

Su, Z., Timmermans, W., Zeng, Y., Schulz, J., John, V. O., Roebeling, R. A., et al. (2018). An overview of european efforts in generating climate data records. *Bulletin of the American Meteorological Society*, *99*(2), 349–359.

Tedersoo, L., Bahram, M., Polme, S., Koljalg, U., Yorou, N. S., Wijesundera, R., et al. (2014). Global diversity and geography of soil fungi. *Science*, *346*(6213), 1078.

Teeri, J. A., & Raven, P. H. (2002). A national ecological observatory network. *Science*, *298*(5600), 1893.

Teruzzi, A., Bolzon, G., Cossarini, G., Lazzari, P., Salon, S., Crise, A., & Solidoro, C. (2019). *Mediterranean Sea biogeochemical reanalysis (CMEMS MED-biogeochemistry) [data set]*. Copernicus Monitoring Environment Marine Service (CMEMS). https://doi.org/10.25423/MEDSEA_REANALYSIS_BIO_006_008

Teruzzi, A., Dobricic, S., Solidoro, C., & Cossarini, G. (2014). A 3-D variational assimilation scheme in coupled transport-biogeochemical models: Forecast of Mediterranean biogeochemical properties. *Journal of Geophysical Research: Oceans*, *119*(1), 200–217. https://doi.org/10.1002/2013JC009277

Terzic, E., Lazzari, P., Organelli, E., Solidoro, C., Salon, S., D'Ortenzio, F., & Conan, P. (2019). Merging bio-optical data from Biogeochemical-Argo floats and models in marine biogeochemistry. *Biogeosciences*, *16*(12), 2527–2542.
Thirel, G., Salamon, P., Burek, P., & Kalas, M. (2013). Assimilation of MODIS snow cover area data in a distributed hydrological model using the particle filter. *Remote Sensing*, *5*(11), 5825–5850.
Toure, A. M., Reichle, R. H., Forman, B. A., Getirana, A., & De Lannoy, G. J. M. (2018). Assimilation of MODIS snow cover fraction observations into the NASA catchment land surface model. *Remote Sensing*, *10*(2).
Ullah, W., Wang, G. J., Gao, Z. Q., Hagan, D. F. T., & Lou, D. (2018). Comparisons of remote sensing and reanalysis soil moisture products over the Tibetan Plateau, China. *Cold Regions Science and Technology*, *146*, 110–121.
Uppala, S., Simmons, A., Dee, D., Kallberg, P., & Thepaut, J. N. (2008). Atmospheric reanalyses and climate variations. *Climate Variability and Extremes during the Past 100 Years*, *33*, 103–117.
Uppala, S. M., Kallberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D., Fiorino, M., et al. (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, *131*(612), 2961–3012.
Ustin, S. L., & Gamon, J. A. (2010). Remote sensing of plant functional types. *New Phytologist*, *186*(4), 795–816. Retrieved from https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8137.2010.03284.x
Valcke, S. (2013). The OASIS3 coupler: A European climate modelling community software. *Geoscientific Model Development*, *6*(2), 373–388.
Valcke, S., Balaji, V., Craig, A., DeLuca, C., Dunlap, R., Ford, R. W., et al. (2012). Coupling technologies for Earth system modelling. *Geoscientific Model Development*, *5*(6), 1589–1596.
van den Hoogen, J., Geisen, S., Routh, D., Ferris, H., Traunspurger, W., Wardle, D. A., et al. (2019). Soil nematode abundance and functional group composition at a global scale. *Nature*, *572*(7768), 194.
van den Hoogen, J., Geisen, S., Wall, D. H., Wardle, D. A., Traunspurger, W., de Goede, R. G. M., et al. (2020). A global database of soil nematode abundance and functional group composition. *Scientific Data*, *7*(1), 103.
van den Hurk, B., & Viterbo, P. (2003). The Torne-Kalix PILPS 2(e) experiment as a test bed for modifications to the ECMWF land surface scheme. *Global and Planetary Change*, *38*(1–2), 165–173.
van der Schalie, R., Kerr, Y. H., Wignerond, J. P., Rodriguez-Fernandez, N. J., Al-Yaari, A., & de Jeu, R. A. M. (2016). Global SMOS soil moisture retrievals from the land parameter retrieval model. *International Journal of Applied Earth Observation and Geoinformation*, *45*, 125–134.
van der Tol, C., Verhoef, W., Timmermans, J., Verhoef, A., & Su, Z. (2009). An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. *Biogeosciences*, *6*(12), 3109–3129. Retrieved from https://bg.copernicus.org/articles/6/3109/2009/
van Dijk, A. I. J. M., Renzullo, L. J., Wada, Y., & Tregoning, P. (2014). A global water cycle reanalysis (2003-2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble. *Hydrology and Earth System Sciences*, *18*(8), 2955–2973.
Verdy, A., & Mazloff, M. R. (2017). A data assimilating model for estimating southern Ocean biogeochemistry. *Journal of Geophysical Research: Oceans*, *122*(9), 6968–6988.
Vereecken, H., Weihermuller, L., Assouline, S., Simunek, J., Verhoef, A., Herbst, M., et al. (2019). Infiltration from the Pedon to global grid scales: An overview and outlook for land surface modeling. *Vadose Zone Journal*, *18*(1), 1–53.
Viskari, T., Hardiman, B., Desai, A. R., & Dietze, M. C. (2015). Model-data assimilation of multiple phenological observations to constrain and predict leaf area index. *Ecological Applications*, *25*(2), 546–558.
Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (1997). Human domination of Earth's ecosystems. *Science*, *277*(5325), 494–499.
von Schuckmann, K., Le Traon, P.-Y., Smith, N., Pascual, A., Djavidnia, S., Gattuso, J.-P., et al. (2020). Copernicus marine service ocean state report, Issue 4. *Journal of Operational Oceanography*, *13*(sup1), S1–S172. https://doi.org/10.1080/1755876X.2020.1785097
Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., & Schoups, G. (2013). Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. *Advances in Water Resources*, *51*, 457–478.
Wagner, W., Dorigo, W., Jeu, R. D., Fernandez, D., Benveniste, J., Haas, E., & Ertl, M. (2012). Fusion of active and passive microwave observations to create an Essential Climate Variable data record on soil moisture. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, (1–7), 315–321.
Wahl, S., Bollmeyer, C., Crewell, S., Figura, C., Friederichs, P., Hense, A., et al. (2017). A novel convective-scale regional reanalysis COSMO-REA2: Improving the representation of precipitation. *Meteorologische Zeitschrift*, *26*(4), 345–361.
Wang, B., Fennel, K., Yu, L. Q., & Gordon, C. (2020). Assessing the value of biogeochemical Argo profiles versus ocean color observations for biogeochemical model optimization in the Gulf of Mexico. *Biogeosciences*, *17*(15), 4059–4074.
Wang, Y. W., Leng, P., Peng, J., Marzahn, P., & Ludwig, R. (2021). Global assessments of two blended microwave soil moisture products CCI and SMOPS with in-situ measurements and reanalysis data. *International Journal of Applied Earth Observation and Geoinformation*, *94*. 102234
Weintraub, S. R., Flores, A. N., Wieder, W. R., Sihi, D., Cagnarini, C., Goncalves, D. R. P., et al. (2019). Leveraging environmental research and observation networks to advance soil carbon science. *Journal of Geophysical Research: Biogeosciences*, *124*(5), 1047–1055.
Weir, B., Miller, R. N., & Spitz, Y. H. (2013). A potential implicit particle method for high-dimensional systems. *Nonlinear Processes in Geophysics*, *20*(6), 1047–1060.
While, J., Totterdell, I., & Martin, M. (2012). Assimilation of pCO(2) data into a global coupled physical-biogeochemical ocean model. *Journal of Geophysical Research*, *117*, https://doi.org/10.1029/2010JC006815
Whitaker, J. S., Compo, G. P., Wei, X., & Hamill, T. M. (2004). Reanalysis without radiosondes using ensemble data assimilation. *Monthly Weather Review*, *132*(5), 1190–1200.
Wieder, W. R., Lawrence, D. M., Fisher, R. A., Bonan, G. B., Cheng, S. J., Goodale, C. L., et al. (2019). Beyond static benchmarking: Using experimental manipulations to evaluate land model assumptions. *Global Biogeochemical Cycles*, *33*(10), 1289–1309. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GB006141
Wilkinson, D. P., Golding, N., Guillera-Arroita, G., Tingley, R., & McCarthy, M. A. (2019). A comparison of joint species distribution models for presence-absence data. *Methods in Ecology and Evolution*, *10*(2), 198–211.
Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*.
Williams, A. P., Cook, E. R., Smerdon, J. E., Cook, B. I., Abatzoglou, J. T., Bolles, K., et al. (2020). Large contribution from anthropogenic warming to an emerging North American megadrought. *Science*, *368*(6488), 314–318. Retrieved from https://science.sciencemag.org/content/sci/368/6488/314.full.pdf

Williams, D. N., Ananthakrishnan, R., Bernholdt, D. E., Bharathi, S., Brown, D., Chen, M., et al. (2009). The Earth system GRID enabling access to multimodel climate simulation data. *Bulletin of the American Meteorological Society*, *90*(2), 195–205.

Williams, M., Schwarz, P. A., Law, B. E., Irvine, J., & Kurpius, M. R. (2005). An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biology*, *11*(1), 89–105.

Wood, S. L. R., Jones, S. K., Johnson, J. A., Brauman, K. A., Chaplin-Kramer, R., Fremier, A., et al. (2018). Distilling the role of ecosystem services in the sustainable development goals. *Ecosystem Services*, *29*, 70–82.

Wuest, R. O., Zimmermann, N. E., Zurell, D., Exander, J. M. A., Fritz, S. A., Hof, C., et al. (2020). Macroecology in the age of big data - Where to go from here? *Journal of Biogeography*, *47*(1), 1–12.

Xiao, Y. J., & Friedrichs, M. A. M. (2014). The assimilation of satellite-derived data into a one-dimensional lower trophic level marine ecosystem model. *Journal of Geophysical Research: Oceans*, *119*(4), 2691–2712.

Xie, J. P., Bertino, L., Counillon, F., Lisaeter, K. A., & Sakov, P. (2017). Quality assessment of the TOPAZ4 reanalysis in the Arctic over the period 1991-2013. *Ocean Science*, *13*(1), 123–144.

Xie, X. Y., Li, A. N., Jin, H. A., Tan, J. B., Wang, C. B., Lei, G. B., et al. (2019). Assessment of five satellite-derived LAI datasets for GPP estimations through ecosystem models. *The Science of the Total Environment*, *690*, 1120–1130.

Xue, Y., Balmaseda, M. A., Boyer, T., Ferry, N., Good, S., Ishikawa, I., et al. (2012). A Comparative analysis of upper-ocean heat content variability from an ensemble of operational ocean reanalyses. *Journal of Climate*, *25*(20), 6905–6929.

Yang, S. H., Li, R., Wu, T. H., Hu, G. J., Xiao, Y., Du, Y. Z., et al. (2020). Evaluation of reanalysis soil temperature and soil moisture products in permafrost regions on the Qinghai-Tibetan Plateau. *Geoderma*, *377*, 114583.

Yu, L. Q., Fennel, K., Bertino, L., El Gharamti, M., & Thompson, K. R. (2018). Insights on multivariate updates of physical and biogeochemical ocean variables using an ensemble Kalman filter and an idealized model of upwelling. *Ocean Modelling*, *126*, 13–28.

Zaitchik, B. F., Rodell, M., & Reichle, R. H. (2008). Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi River basin. *Journal of Hydrometeorology*, *9*(3), 535–548.

Zhang, Y., Pan, M., Sheffield, J., Siemann, A. L., Fisher, C. K., Liang, M. L., et al. (2018). A Climate Data Record (CDR) for the global terrestrial water budget: 1984-2010. *Hydrology and Earth System Sciences*, *22*(1), 241–263.

Zhang, Y., Xiao, X. M., Wu, X. C., Zhou, S., Zhang, G. L., Qin, Y. W., & Dong, J. W. (2017). Data Descriptor: A global moderate resolution dataset of gross primary production of vegetation for 2000-2016. *Scientific Data*, *4*.

Zhao, S., Pederson, N., D'Orangeville, L., HilleRisLambers, J., Boose, E., Penone, C., et al. (2019). The International Tree-Ring Data Bank (ITRDB) revisited: Data availability and global ecological representativity. *Journal of Biogeography*, *46*(2), 355–368. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/jbi.13488

Zhao, Y., Chen, X. Q., Smallman, T. L., Flack-Prain, S., Milodowski, D. T., & Williams, M. (2020). Characterizing the error and bias of remotely sensed LAI Products: An Example for Tropical and Subtropical Evergreen Forests in South China. *Remote Sensing*, *12*(19).

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., & Mayer, M. (2019). The ECMWF operational ensemble reanalysis-analysis system for ocean and sea ice: A description of the system and assessment. *Ocean Science*, *15*(3), 779–808.