



## Deep learning and process understanding for data-driven Earth system science

### Postprint version

Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung,  
Joachim Denzler, Nuno Carvalhais & Prabhat

**Published in: Nature**

**Reference:** Reichstein, M., Camps-Valls, G., Stevens, B. et al. Deep learning and process understanding for data-driven Earth system science. *Nature* 566, 195–204 (2019). <https://doi.org/10.1038/s41586-019-0912-1>

**Web link:** <https://www.nature.com/articles/s41586-019-0912-1>



This paper has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 647423

# Deep learning and process understanding for data-driven Earth System Science

Markus Reichstein<sup>1,2)</sup>, Gustau Camps-Valls<sup>3)</sup>, Bjorn Stevens<sup>4)</sup>, Martin Jung<sup>1)</sup>, Joachim Denzler<sup>2),5)</sup>, Nuno Carvalhais<sup>1,6)</sup>, Mr Prabhat<sup>7)</sup>

1. Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Jena, Germany
2. Michael-Stifel-Center Jena for Data-driven and Simulation Science, Jena, Germany
3. Image Processing Laboratory (IPL), University of València, Spain
4. Max Planck Institute for Meteorology, Hamburg, Germany
5. Computer Vision Group, Computer Science, Friedrich Schiller University, Jena, Germany
6. CENSE, Departamento de Ciências e Engenharia do Ambiente, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, Caparica, Portugal
7. National Energy Research Supercomputing Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA

## **Summary paragraph**

Machine learning approaches are increasingly used to extract patterns and insights from the exploding universe of geospatial data, but current approaches may not be an optimal approach when system behavior is dominated by spatial or temporal context. Rather than amending classical machine learning, however, we argue that these contextual cues should be at the core of a modified approach – termed deep learning – to extract novel understanding and predictive ability for topics such as seasonal forecasting and modeling of long-range spatial connections across multiple time-scales. A critical further step will be a hybrid modeling approach coupling physical processes with deep learning versatility.

## **1. Introduction**

Humans have always been striving to predict and understand the world, and the ability to make better predictions has given competitive advantages in diverse contexts (e.g., weather, diseases, or more recently financial markets). Yet the tools for prediction have substantially changed over time, from ancient Greek philosophical reasoning to non-scientific medieval methods like soothsaying, toward modern scientific discourse, which has come to include hypothesis testing, theory development and computer modelling underpinned by statistical and/or physical relationships, i.e., laws<sup>1</sup>. A success story in the geosciences is weather prediction, which has greatly improved through integration of better theory, increased

37 computational power, and established observational systems which allow for the assimilation of  
38 large amounts of data into the modeling system<sup>2</sup>. Nevertheless, we can only accurately predict  
39 the evolution of the weather on a time-scale of days, not months. Seasonal meteorological  
40 predictions, forecasting extreme events such as flooding or fire, and long-term climate  
41 projections are still major challenges. This is especially true for predicting dynamics in the  
42 biosphere, which is dominated by biologically mediated processes such as growth, or  
43 reproduction, and strongly controlled by the seemingly stochastic disturbances such as fires and  
44 landslides. Such problems have been rather resistant to progress in the past decades<sup>3</sup>.

45 At the same time, a deluge of Earth system data has become available, with storage volumes  
46 already well beyond dozens of petabytes and with rapidly increasing transmission rates beyond  
47 hundreds of terabytes per day<sup>4</sup>. These data come from a plethora of sensors measuring states,  
48 fluxes, and intensive or time/space integrated variables, and representing fifteen or more orders  
49 of temporal and spatial magnitude. They include remote sensing from meters to hundreds  
50 kilometers above the Earth as well as in-situ observations (increasingly from autonomous  
51 sensors) at and below the surface and in the atmosphere, many of which are further being  
52 complemented by citizen science observations. Model simulation output adds to this deluge; the  
53 CMIP-5 dataset (Climate Model Intercomparison Project), used extensively by the scientific  
54 community for scientific groundwork towards periodic climate assessments, is over 3PB in size,  
55 and the next generation, CMIP-6, is estimated to reach up to 30PB<sup>5</sup>. While not observations, the  
56 model data share many of the challenges and statistical properties of observational data,  
57 including many forms of uncertainty. In summary, Earth System data are exemplary of all four of  
58 the “four V’s” of Big Data: volume, velocity, variety, and veracity (Figure 1). One key challenge is  
59 to extract interpretable information and knowledge from this Big Data, possibly in near-real time  
60 and integrating between disciplines.

61 Taken together, our ability to collect and create data far outpaces our ability to sensibly  
62 assimilate it, let alone understand it. Predictive ability in the last few decades has not increased  
63 apace with data availability. To get the most out of the explosive growth and diversity of Earth  
64 system data, we face two major tasks in the coming years: 1) extracting knowledge from the  
65 data deluge, and 2) deriving models which learn maximally from data, beyond traditional data  
66 assimilation approaches, while still respecting our evolving understanding of nature’s laws.

67 The combination of unprecedented data sources, increased computational power, and the  
68 recent advances in statistical modeling and machine learning offer exciting new opportunities for  
69 expanding our knowledge about the Earth system from data. In particular, many tools are  
70 available from the fields of machine learning and artificial intelligence, but they need to be  
71 further developed and adapted to geo-scientific analysis. Earth system science offers new  
72 opportunities, challenges and methodological demands, in particular for recent research lines  
73 focusing on spatio-temporal context and uncertainties (see Glossary).

74 *[Place Glossary around here]*

75 In the following sections we review the development of machine learning in the geoscientific  
76 context, and highlight how deep learning, i.e. the automatic extraction of abstract (spatio-  
77 temporal) features, has the potential to overcome many of the limitations that have, until now,  
78 hindered a more wide-spread adoption of machine learning. We further lay out the most  
79 promising but also challenging approaches in combining machine learning with physical  
80 modelling.

## 81 **2. State-of-the-art in geoscientific machine learning**

82 Machine learning is now a successful part of several research-driven and operational  
83 geoscientific processing schemes, addressing the atmosphere, the land surface and the ocean,  
84 but has co-evolved with data availability over the last decade. Early landmarks in classification  
85 of land cover and clouds emerged almost 30 years ago through the coincidence of high-  
86 resolution satellite data and the first revival of neural networks<sup>6,7</sup>. Most major machine learning  
87 methodological development (e.g. kernel methods or Random forests) has subsequently been  
88 applied to geoscience and remote sensing problems, often when data suitable for pertinent  
89 methods became available<sup>8</sup>. Thus, machine learning has become a universal approach in geo-  
90 scientific classification, and change and anomaly detection problems<sup>9,10-12</sup>. In the last few years,  
91 the field has begun to use deep learning to better exploit spatial and temporal structure in the  
92 data, features that would normally be problematic for traditional machine learning (e.g. Table 1,  
93 and next section).

94 Another class of problems where machine learning has been successful is regression problems.  
95 An example is soil mapping, where measurements of soil properties and covariates exist at  
96 points sparsely distributed in space, and where a Random Forest, a popular and efficient

97 machine learning approach, is used to predict spatially dense estimates of soil properties or soil  
98 types<sup>13,14</sup>. In the last decade, machine learning has attained outstanding results in the  
99 regression estimation of bio-geo-physical parameters from remotely sensed reflectances at local  
100 and global scales<sup>15,16,17</sup>. These approaches emphasize spatial prediction, i.e. prediction of  
101 properties which are relatively static over the observational time period.

102 Yet, what makes the Earth System interesting is that it is not static, but dynamic. Machine  
103 learning regression techniques have also been utilized to study these dynamics by mapping  
104 temporally varying features onto temporally varying target variables in land, ocean and  
105 atmosphere domains. Since variables such as land- or ocean-atmosphere carbon uptake  
106 cannot be observed everywhere, one challenge has been to infer continental or global estimates  
107 from point observations, by building models, which relate climate and remote sensing co-  
108 variates to the target variables. In this context, machine learning methods have proven to be  
109 more powerful and flexible than previous mechanistic or semi-empirical modelling approaches..  
110 For instance an ANN with one hidden layer was able to filter out noise, predict the diurnal and  
111 seasonal variation of CO<sub>2</sub> fluxes, and extract patterns such as an increased respiration in spring  
112 during root growth, which was formerly unquantified and not well represented in carbon cycle  
113 models<sup>18</sup>. Further developments have then allowed for the first time to quantify global terrestrial  
114 photosynthesis and evapotranspiration of water in a purely data-driven way<sup>19,20</sup>. Spatial,  
115 seasonal, interannual or decadal variation of such machine-learning-predicted fluxes are even  
116 being used as important benchmarks for physical land-surface and climate model evaluation<sup>21-</sup>  
117 <sup>24</sup>. Similarly, ocean CO<sub>2</sub> concentrations and fluxes have been mapped spatio-temporally with  
118 neural networks, where classification and regression approaches have been combined, both for  
119 stratifying the data and for prediction<sup>25</sup>. Recently random forests have also been used to predict  
120 spatio-temporally varying precipitation<sup>26</sup>. Overall, we conclude that a diversity of influential  
121 machine learning approaches have already been applied across all the major sub-domains of  
122 Earth system science and are increasingly being integrated into operational schemes and being  
123 used to discover new patterns, advance understanding and evaluate comprehensive physical  
124 models.

125 Notwithstanding the success of machine learning in the geosciences, important caveats and  
126 limitations have hampered a wider adoption and impact of such methods. A few pitfalls such as  
127 the risk of naïve extrapolation, sampling or other data biases, ignorance of confounding factors,  
128 interpretation of statistical association as causal relation, or fundamental flaws in multiple

129 hypothesis testing (“p-fishing”) <sup>27-29</sup> should be avoided by best practices and expert intervention.  
130 More fundamentally, there are inherent limitations of currently-applied machine learning  
131 approaches. It is in this realm that the techniques of deep learning promise breakthroughs, as  
132 we explain in the paragraphs below.

133 Classical machine learning approaches benefit from domain-specific, hand-crafted features to  
134 account for dependencies in time or space (e.g. cumulative precipitation derived from a daily  
135 time series), but rarely exploit spatio-temporal dependencies exhaustively. For instance, in  
136 ocean-atmosphere or land-atmosphere CO<sub>2</sub> flux prediction<sup>19,25</sup>, mapping of instantaneous, local  
137 environmental conditions (e.g. radiation, temperature, humidity) to instantaneous fluxes is  
138 performed. In reality, processes at a certain point in time and space are almost always  
139 additionally affected by the state of the system, which is often not well observed and thus not  
140 available as a predictor. However, previous time steps and neighboring grid cells contain hidden  
141 information on the state of the system (e.g. a long period without rain-fall combined with  
142 sustained sunny days implies a drought). One example where both, spatial and temporal  
143 context are highly relevant, is the prediction of fire occurrence and characteristics such as burnt  
144 area and trace gas emissions. Fire occurrence and spread depends not only on instantaneous  
145 climatic drivers and sources of ignition (e.g. humans, lightning, or both) but also on state  
146 variables, such as the state and amount of available fuel<sup>3</sup>. Fire spread and thus the burnt area  
147 depends not only on the local conditions of each pixel but also on the spatial arrangement and  
148 connectivity of fuel, its moisture, terrain properties, and of course wind speed and direction.  
149 Similarly, classifying a certain atmospheric situation as a hurricane or extratropical storm  
150 requires knowledge of the spatial context such as size and shape of a geometry constituted by  
151 pixels, their values, and their topology. For instance, detecting symmetric outflow and a visible  
152 ‘eye’ is important for detecting hurricanes and assessing their strength which cannot be  
153 determined alone by localized, single pixel values.

154 Certainly, temporally dynamic properties (“memory effects”) can be represented by hand-  
155 designed and domain-specific features in machine learning. Examples are cumulative sums of  
156 daily temperature, which are used to predict phenological phases of vegetation, and the  
157 standardized precipitation index (SPI<sup>30</sup>), which summarizes precipitation anomalies over the last  
158 months as a meteorological indicator of drought states. Very often, these approaches only  
159 consider memory in a single variable, ignoring interactive effects of several variables, although  
160 exceptions exist <sup>22,31</sup>.

161 Machine learning can also use hand-designed features, such as terrain shape and  
162 topographical or texture features from satellite images, to incorporate spatial context<sup>6</sup>. This is  
163 analogous to earlier approaches in computer vision where objects were often characterized by a  
164 set of features describing edges, textures, shapes and colors. Such features were then fed into  
165 a standard machine learning for localization, classification or detection of objects in images.  
166 Similar approaches have been followed for decades in remote sensing image classification<sup>8-10</sup>.  
167 Hand-designed features can be seen both as an advantage (control of the explanatory drivers)  
168 and as a disadvantage (tedious, *ad hoc* process, likely non-optimal), but certainly the concern of  
169 a restricted, and subjective choice of features rather than an extensive and generic approach  
170 remains a valid and important one. New developments in deep learning, however, no longer  
171 limit us to such approaches.

### 172 **3. Deep-learning opportunities in Earth system science**

173 Deep learning has achieved notable success in modelling ordered sequences and data with  
174 spatial context in the fields of computer vision, speech recognition and control systems<sup>32</sup>, as  
175 well as in related scientific fields in physics<sup>33-35</sup>, chemistry<sup>36</sup> and biology<sup>37</sup> (see also ref<sup>38</sup>).  
176 Applications to problems in geosciences are in their infancy, but across the key problems  
177 (classification, anomaly detection, regression, space- or time dependent state prediction) there  
178 are promising examples arising (Table 1, Supplementary Box 1)<sup>39,40</sup>. Two recent studies  
179 demonstrate the application of deep learning to the problem of extreme weather, for instance  
180 hurricane, detection<sup>41,42</sup> – already mentioned as a problematic question for traditional machine  
181 learning” They report success in applying deep-learning architectures to objectively extract  
182 spatial features define and classify extreme situations (e.g. storms, atmospheric rivers) in  
183 numerical weather prediction model output. Such approach enables rapid detection of such  
184 events and forecast simulations without using either subjective human annotation or methods  
185 that rely on predefined somewhat arbitrary thresholds for wind speed or other variables. In  
186 particular, such approach uses the information in the spatial shape of respective events such as  
187 the typical spiral for hurricanes. Similarly, for classification of urban areas the automatic  
188 extraction of multi-scale features from remote sensing data strongly improved the classification  
189 accuracy to almost always greater than 95%<sup>43</sup>.

190 While deep learning approaches have classically been divided into spatial learning (e.g.  
191 convolutional neural networks for object classification) and sequence learning (e.g. speech

192 recognition), there is a growing interest in blending these two perspectives. A prototypical  
193 example is video and motion prediction<sup>44,45</sup>, which is strikingly similar to many dynamic  
194 geoscience problems. Here we are faced with time-evolving multi-dimensional structures, such  
195 as organized precipitating convection which dominates patterns of tropical rainfall, vegetation  
196 states which influence the flow of carbon and evapotranspiration. Studies are beginning to apply  
197 combined convolutional-recurrent approaches to geoscientific problems such as precipitation  
198 nowcasting (Table 1)<sup>46</sup>. Modelling atmospheric and ocean transport, fire spread, soil movements  
199 or vegetation dynamics are other examples where spatio-temporal dynamics are important, but  
200 which have yet to benefit from a concerted effort to apply these new approaches.

201 In short, the similarities between the types of data addressed with classical deep learning  
202 applications and geoscientific data make a compelling argument for the integration of deep  
203 learning into geosciences (Figure 2): Images are analogous to two-dimensional data fields  
204 containing particular variables in analogy to color-triplets (RGB values) in photographs, while  
205 videos can be likened to a sequence of images and hence of 2D fields that evolve in time.  
206 Similarly, natural language and speech signals share the same multiresolution characteristics of  
207 dynamic time-series of Earth system variables. Furthermore, classification, regression, anomaly  
208 detection, and dynamic modeling are typical problems in both computer vision and geosciences.

#### 209 **4. Deep-learning challenges in Earth system science**

210 The similarities between classical deep learning applications and geoscience applications  
211 outlined above are striking. Yet, numerous differences exist. For example, while classical  
212 computer vision applications deal with photos which have three channels (red, green, blue)  
213 hyperspectral satellite images extend to hundreds of spectral channels well beyond the visible  
214 range, which often induce different statistical properties to those of natural images. This  
215 includes spatial dependence and interdependence of variables violating the important  
216 assumption of identically, independent distributed data. Additionally, integrating multi-sensor  
217 data is not trivial since different sensors exhibit different imaging geometries, spatial and  
218 temporal resolution, physical meaning, content and statistics. Sequences of (multi-sensor)  
219 satellite observations also come with diverse noise sources, uncertainty levels, missing data  
220 and (often systematic) gaps (due to the presence of clouds or snow, distortions in the  
221 acquisition, storage and transmission, etc.).



222 In addition, spectral, spatial, and temporal dimensionalities raise computational challenges. The  
223 data volume is increasing geometrically and soon it will be necessary to deal with Petabytes/day  
224 globally. Currently, the biggest meteorological agencies have to process Terabytes per day in  
225 near real time. often at very high precision (32-bit, 64-bit). Further, while typical computer vision  
226 applications have worked with image sizes of 512 x 512 pixels, a moderate resolution (ca. 1km)  
227 global field has sizes of approximately 40000 x 20000 pixels, i.e. three orders of magnitude  
228 more.

229 Last but not least, unlike the ImageNet benchmark (a data base of images with labels, e.g. “cat”  
230 or “dog”<sup>47</sup>) in the computer vision community, large, labeled geoscientific datasets do not always  
231 exist in geo-science, not only due to the sizes of the datasets involved, but also due to the  
232 conceptual difficulty in labeling data sets, e.g. determining “it’s a cat” vs “it’s a drought”, given  
233 that the second label is contingent on intensity and extent and can change according to  
234 methods, and there are not enough labeled cases for training. These aspects raise the  
235 challenge of working with a limited training set. More generally, geo-scientific problems are often  
236 underconstrained, leading to the possibility of models thought to be of high quality, which  
237 perform well in training and even test data sets, but deviate strongly for situations and data  
238 outside their valid domain (extrapolation problem), which is even true for complex physical Earth  
239 system models<sup>48</sup>. Overall, we identify at least five major challenges and avenues for the  
240 successful adoption of deep learning approaches in the geosciences:

- 241 1. Interpretability: Improving predictive accuracy is important but insufficient. Certainly,  
242 interpretability and understanding are crucial in this arena, including visualization of the  
243 results for analysis by humans. Interpretability has been identified as a potential weakness  
244 of deep neural networks, and achieving it is a current focus in deep learning<sup>49</sup>. The field is  
245 still far from achieving self-explanatory models, and from causal discovery from  
246 observational data<sup>50,51</sup>. Yet, we should note that, given their complexity, also modern Earth  
247 system models are in practice often not easily traceable back to their assumptions, limiting  
248 their interpretability as well.
- 249 2. Physical consistency: Deep learning models can fit observations very well, but predictions  
250 may be physically inconsistent or implausible, e.g. owing to extrapolation or observational  
251 biases. Integration of domain knowledge and achievement of physical consistency by  
252 teaching models about the governing physical rules of the Earth system can provide very  
253 strong theoretical constraints on top of the observational ones.

- 254 3. Complex and uncertain data: New deep learning methods are needed to cope with complex  
255 statistics, multiple outputs, different noise sources and high dimensional spaces. New  
256 network topologies that not only exploit local neighborhood (even at different scales), but  
257 also long-range relationships (e.g., for teleconnections) are urgently needed, but the exact  
258 cause-effect relations between variables are even not clear in advance and need to be  
259 discovered. Modelling uncertainties will be certainly an important aspect and will require to  
260 integrate concepts from Bayesian/probabilistic inference, which are directly addressing that  
261 (Glossary and <sup>52</sup>).
- 262 4. Limited labels: Methods need to be further developed which can learn from few labelled  
263 examples, by utilizing the information in related unlabeled observations, so-called  
264 unsupervised density modeling, feature extraction and semi-supervised learning<sup>53</sup> (cf.  
265 glossary).
- 266 5. Computational demand: There is a huge technical challenge regarding the high  
267 computational cost of current geoscience problems - good examples to address this  
268 includes Google Earth Engine, which allowed solving real problems from deforestation<sup>54</sup> to  
269 lake<sup>55</sup> monitoring, yet still without deep learning application.

270 By addressing these challenges, deep learning could make an even bigger difference in the  
271 geosciences in comparison to classical computer vision, because in computer vision hand  
272 crafted features are derived from a clear understanding of the world (existence of surfaces,  
273 boundaries between objects, etc.), the mapping from the world to images, and assumptions  
274 about the (visual) appearance of world points (surface points, the state in 3D) on 2D images.  
275 Assumptions for successful processing include the assumption of Lambertian surfaces (i.e.  
276 intensity does not depend on the angle between surface and light source) which results in the  
277 classical assumption of constant intensity of the observation of a 3D point over time. In addition,  
278 changes in the world (the motion of objects) are in most cases modeled as rigid transformations,  
279 or non-rigid transformations that arise from physical assumptions and that are only valid locally  
280 (like in registration of brain structures, before and after removal of a tumor). Even complex  
281 problems in computer vision have been solved by hand-crafted features that reflect the  
282 assumptions and expectations arising from common world knowledge. In geoscience and  
283 climate science, such global, general assumptions are still partly missing. In fact, these  
284 assumptions and expectations are exactly the models we are looking for! All problems, from  
285 segmentation in remote sensing images to regression analysis of certain variables, have certain

286 assumptions that are known to be valid or at least good approximations. Yet, the less processes  
287 are understood, the fewer high-quality hand-crafted features for modeling are expected to exist.  
288 Thus, deep learning methods, particularly since they find a good representation from data,  
289 represent an opportunity to tackle geoscience and climate research problems.

290 The most promising near-future applications include nowcasting, (i.e. prediction of the very near  
291 future, up to two hours in meteorology) and forecasting applications, anomaly detection and  
292 classification based on spatial and temporal context information (see examples in Table 1). A  
293 longer-term vision includes data driven seasonal forecasting, modelling of spatial long-range  
294 correlations across multiple time-scales, modelling spatial dynamics where spatial context plays  
295 an important role (e.g. fires), and detecting teleconnections and connections between variables  
296 that a human may not have thought about.

297 Overall, we infer that deep learning will soon be the leading method for classifying and  
298 predicting space-time structures in the geosciences. More challenging is to gain understanding  
299 in addition to optimal prediction, and to achieve models that have maximally learned from data,  
300 while still respecting and taking advantage of the physical and biological knowledge. One  
301 promising but largely uncharted approach to achieving this goal is the integration of machine  
302 learning with physical modelling, which we explore in the following section.

303

## 304 **5 Integration with physical modelling**

305 Historically, physical modelling and machine learning have been often treated as “two different  
306 worlds” with very different scientific paradigms (theory-driven versus data-driven). Yet, in fact  
307 these approaches are complementary, with physical approaches in principle being directly  
308 interpretable and offering the potential of extrapolation beyond observed conditions, while data-  
309 driven approaches are highly flexible in adapting to the data and are amenable to finding  
310 unexpected patterns (surprises). The synergy between the two approaches has been gaining  
311 attention<sup>56-58</sup>, expressed in benchmarking initiatives<sup>59,60</sup> and in concepts such as emergent  
312 constraints<sup>27,61,62</sup>.

313 Here, we argue that advances in machine learning and in observational and simulation  
314 capabilities within Earth sciences offer an opportunity to more intensively integrate simulation  
315 and data science approaches in multiple fashions. From a systems modelling point of view there

316 are five points of potential synergy (Figure 3) [the numbers in the following list correspond to the  
317 circles in the figure]:

318 1) Improving parameterizations (Fig. 3, linkage 1). Physical models require parameters but  
319 many of those cannot be easily derived from first principles. Here, machine learning can learn  
320 parameterizations to optimally describe the ground truth which can be observed or generated  
321 from detailed and high-resolution models through first principles. For example, instead of  
322 assigning parameters of the vegetation in an Earth system model to plant functional types (a  
323 common *ad hoc* decision in most global land surface models), one can allow these  
324 parameterizations to be learned from appropriate sets of statistical covariates, allowing them to  
325 be more dynamic, interdependent and contextual. A prototypical approach has been taken  
326 already in hydrology where the mapping of environmental variables (e.g. precipitation, surface  
327 slope) to catchment parameters (e.g. mean, minimum, maximum streamflow) has been learned  
328 from a few thousands catchments and applied globally to feed hydrological models<sup>63</sup>. Another  
329 example from global atmospheric modelling is learning the effective coarse-scale physical  
330 parameters of precipitating convection (e.g. the fraction of water that is precipitating out of a  
331 cloud during convection) from data or high-resolution models<sup>64,65</sup>. (the high-resolution models are  
332 too expensive to run, which is why coarse-scale parametrizations are needed). These learned  
333 parametrizations could lead to better representations of tropical convection<sup>66,67</sup>.

334 2) Replacing a “physical” sub-model with a machine learning model (Fig. 3, linkage 2). If  
335 formulations of a submodel are of semi-empirical nature where the functional form has little  
336 theoretical basis (e.g. biological processes), this submodel can be replaced by a machine  
337 learning model if a sufficient number of observations are available. This leads to a *hybrid* model,  
338 which combines the strengths of physical modeling (theoretical foundations, interpretable  
339 compartments) and machine learning (data-adaptiveness). For example, we could couple well  
340 established physical (differential) equations of diffusion for transport of water in plants with  
341 machine learning for the poorly understood biological regulation of water transport conductance.  
342 This results in a more “physical model” that obeys accepted conservation of mass and energy  
343 laws, but the regulation (biological) is flexible and learned from data. Such principle has recently  
344 been taken to efficiently model motion of water in the ocean and specifically predict sea surface  
345 temperatures. Here, the motion field was learned via a deep neural network, and then used to  
346 update the heat content and temperatures via physically modelling the movement implied by the  
347 motion field<sup>68</sup>. Also a number of atmospheric scientists have begun experimenting with related

348 approaches to circumvent long-standing biases in physically based parameterizations of  
349 atmospheric convection<sup>65,69</sup>.

350 The problem may become more complicated if physical model and machine learning  
351 parameters are to be estimated simultaneously while maintaining interpretability, especially  
352 when several sub-models are replaced with machine learning approaches. In the field of  
353 chemistry this approach has been used in calibration exercises and to describe changes in  
354 unknown kinetic rates while maintaining mass balance in biochemical reactors modeling<sup>70</sup>,  
355 which, albeit less complex, bears many similarities to hydrological and biogeochemical  
356 modelling.

357 3) Analysis of model-observation mismatch (Fig. 3, linkage 3): Deviations of a physical model  
358 from observations can be perceived as imperfect knowledge causing model error, assuming no  
359 observational biases. Machine learning can help to identify, visualize and understand the  
360 patterns of model error, which allows also to correct model outputs accordingly. For example,  
361 machine learning can extract patterns from data automatically and identify those which are not  
362 explicitly represented in the physical model. This approach helps improving the physical model  
363 and theory. In practice, it can also serve to correct model bias of dynamic variables, or it can  
364 facilitate improved downscaling to finer spatial scales compared to tedious and ad hoc hand-  
365 designed approaches<sup>71,72</sup>.

366 4) Constraining sub-models (Fig. 3, linkage 4). One can drive a submodel with the output from a  
367 machine learning algorithm, instead of another (potentially biased) submodel in an offline  
368 simulation. This helps in disentangling model error originating from the submodule of interest  
369 from errors of coupled submodules. As a consequence, this simplifies and reduces biases and  
370 uncertainties in model parameter calibration or the assimilation of observed system state  
371 variables.

372 5) Surrogate modeling or emulation: Emulation of the full (or specific parts of) a physical model  
373 can be useful for computational efficiency and tractability reasons. Machine learning emulators  
374 once trained can achieve orders of magnitude faster simulations than the original physical  
375 model without sacrificing significant accuracy. This allows for fast sensitivity analysis, model  
376 parameter calibration, and derivation of confidence intervals for the estimates. For example,  
377 machine learning emulators are used to replace computationally expensive, physics-based

378 radiative transfer models (RTMs) of the interactions between radiation, vegetation and  
379 atmosphere<sup>57,73,74</sup> which are critical for the interpretation and assimilation of land surface remote  
380 sensing in models. Emulators are also used in dynamic modelling, where states are evolving,  
381 e.g. in climate modeling<sup>75</sup> and more recently explored in vegetation dynamic models<sup>76</sup>. Further,  
382 given the complexity of physical models, emulation challenges are very good test beds to  
383 explore the potential of machine learning and deep learning approaches to extrapolate outside  
384 the ranges of training conditions.

385 Some of the concepts in Figure 3 have already been adopted in a broad sense. For instance,  
386 point 3) relates to model benchmarking and statistical downscaling and model output  
387 statistics<sup>77,78</sup>. Here we argue that adopting a deep-learning approach will strongly improve the  
388 use of spatio-temporal context information for the modification of model output. Emulation (5)  
389 has been widely adopted in several branches of engineering and geosciences, mainly for the  
390 sake of efficient modelling, but tractability issues have not yet been explored in depth. Other  
391 paths, such as the hybrid modelling (Fig. 3, link 2), appear to be much less explored.  
392 Conceptually the hybrid approaches discussed before can be interpreted as deepening and  
393 “physicizing” a neural network (Figure 4), where the physical model comes on top of a neural  
394 network layers (see examples Fig. 4b-c). It contrasts the reverse approach discussed above  
395 where physical model output is produced and then corrected using additional layers of machine  
396 learning approaches. We believe that it is worthwhile pursuing both avenues of integrating  
397 physical modelling and machine learning.

398 Figure 3 started from a system-modelling view and seeks to integrate machine learning. As an  
399 alternative perspective system knowledge can be integrated into a machine learning framework.  
400 This may include respective design of the network architecture<sup>36,79</sup>, physical constraints in the  
401 cost function for optimization<sup>58</sup>, or expansion of the training data set for under-sampled domains  
402 (i.e. physically based data augmentation)<sup>80</sup>. For instance, while usually a so-called cost-function  
403 like ordinary least squares penalizes model-data mismatch, it can be modified to also avoid  
404 physically implausible predictions for lake temperature modelling<sup>58</sup>. The integration of physics  
405 and machine learning models may not only achieve improved performance and generalizations  
406 but, perhaps more importantly, incorporates consistency and credibility of the machine learning  
407 models. As a by-product, the hybridization has an interesting regularization effect as physics  
408 discards implausible models. Therefore, physics-aware machine learning models should better  
409 combat overfitting, especially in low-to-medium sample sized datasets<sup>81</sup>. This notion is also

410 related to the direction of attaining explainable and interpretable machine learning models  
411 (“explainable AI”<sup>82</sup>), and to combining logic rules with deep neural networks<sup>83</sup>

412 Recent advancements in two fields of methodological approaches have potential in facilitating  
413 the fusion of machine learning and physical models in a sound way: probabilistic  
414 programming<sup>52</sup>, and differentiable programming. Probabilistic programming allows for  
415 accounting of various uncertainty aspects in a formal but flexible way. A proper accounting for  
416 data and model uncertainty along with integration of knowledge by priors and constraints is  
417 critical for optimally combining the data-driven and theory-driven paradigms, including logical  
418 rules as done in statistical relational learning. In addition, error propagation is conceptually  
419 seamless, facilitating well founded uncertainty margins for model output. This capability is  
420 largely missing so far but crucial for scientific purposes, and in particular for management, or  
421 policy decisions. Differentiable programming allows for efficient optimization due to automated  
422 differentiation<sup>84,85</sup>. This greatly helps in making the large, non-linear and complex inversion  
423 problem computationally more tractable, and in addition allows for explicit sensitivity  
424 assessments, thus aiding in interpretability.

425

## 426 **6. Advancing science**

427 There is no doubt and there are numerous examples as discussed in this manuscript, that  
428 modern machine learning methods significantly improve classification and prediction skills. This  
429 alone has great value. Yet, how do they improve fundamental scientific understanding, given  
430 that in particular the outcome of complex statistical models remains hard to grasp? The answer  
431 can be found in the observations which have virtually always been the basis for scientific  
432 progress. The Copernican revolution was possible by precisely observing planetary trajectories  
433 to infer and test the laws governing them. While the general cycle of exploration, hypotheses  
434 generation and testing remains the same, modern data-driven science and machine learning  
435 can extract arbitrarily complex patterns in observational data to challenge complex theories and  
436 Earth system models (Supplementary Fig. 3). For instance spatially explicit global data-driven  
437 machine learning based estimates of photosynthesis, has indicated an overestimation of  
438 photosynthesis in the tropical rainforest by climate models<sup>86</sup>. This mismatch has led scientists  
439 to develop hypotheses that enable a better description of the radiative transfer in vegetation  
440 canopies<sup>23</sup> which has led to better photosynthesis estimates also in other regions, and better  
441 consistency with leaf level observations.. Related data-driven carbon cycle estimates have

442 helped calibrating vegetation models and explain the conundrum of the increasing seasonal  
443 amplitude of the CO<sub>2</sub> concentration in high latitudes<sup>87</sup>, which according to these results is  
444 caused by more vigorous vegetation in the high latitudes. In addition to data-driven theory and  
445 model building, extracted patterns are increasingly being used as a way to explore improved  
446 parameterizations in Earth system models<sup>65,69</sup>, and emulators are increasingly being used as a  
447 basis for model calibration<sup>88</sup>. In other words, the scientific interplay between theory and  
448 observation, of hypothesis generation and theory-driven hypothesis testing will prevail, but the  
449 complexity of hypotheses and tests inferred from data and the pace of this generation are  
450 changing by orders of magnitude, implying unprecedented, qualitative and quantitative progress  
451 of the science of the complex Earth system.

452

## 453 **7. Conclusion**

454

455 Earth sciences face the need to process large and rapidly increasing amounts of data to provide  
456 more accurate, less uncertain, and physically consistent inferences in the form of prediction,  
457 modeling and understanding the complex Earth system. Machine learning in general, and deep  
458 learning in particular, offer promising tools to build new data-driven models for components of  
459 the Earth system and thus for understanding of the Earth. The Earth system specific challenges  
460 shall further stimulate the development of methodologies, where we have four major  
461 recommendations.

462 *Recognition of the particularities of the data:* multi-source, multi-scale, high dimensional,  
463 complex spatial-temporal relations, including non-trivial, and lagged long-distance relationships  
464 (teleconnections) between variables need to be adequately modelled. While the deep learning  
465 approach is well-positioned to address these data challenges, this may stimulate development  
466 of new network architectures, algorithms and approaches, in particular deep-learning  
467 approaches which address both spatial and temporal context at different scales (cf. Figure 4).

468 *Plausibility and interpretability of inferences:* models should not only be accurate but also  
469 credible and aware of the physics governing the Earth system. Wide adoption of machine  
470 learning in the Earth sciences will be facilitated if models become more transparent and  
471 interpretable: their parameters and feature rankings should have a minimal physical



472 interpretation, and the model should be reducible/explainable in a set of rules, descriptors, and  
473 relations.

474 *Uncertainty estimation:* Models should speak about their confidence and credibility. A strong  
475 integration of Bayesian/probabilistic inference will be an avenue to follow here, because they  
476 allow for explicit representation and propagation of uncertainties. In addition, identifying and  
477 treating extrapolation is a priority.

478 *Testing against complex physical models:* the spatial and temporal prediction ability of machine  
479 learning should be at least consistent with the patterns observed in physical models. Thus we  
480 recommend testing the performance of machine learning methods against synthetic data  
481 derived from physical models of the Earth system. For instance, the models in Fig. 4b and c,  
482 which are applied to real data, should be tested across a broad range of dynamics as simulated  
483 by complex physical models. This is of particular relevance in conditions of limited training data  
484 and to assess extrapolation issues.

485 Overall we suggest that future models should integrate process-based and machine learning  
486 approaches. Data-driven machine learning approaches to geo-scientific research will not  
487 replace physical modelling, but strongly complement and enrich it. Specifically, we envision  
488 various synergies between physical and data-driven models, with the ultimate goal of hybrid  
489 modelling approaches: they obey physical laws, feature a conceptualized and thus interpretable  
490 structure, and at the same time are fully data-adaptive where theory is weak. Importantly, the  
491 other way around also holds: machine learning research will benefit from plausible physically  
492 based relationships derived from the natural sciences. Among others, two major Earth system  
493 challenges resistant to past progress, the parameterization of atmospheric convection and the  
494 description of spatio-temporal dependency of ecosystems on climate and interacting geo-  
495 factors, are open to be addressed with the approaches discussed here.

#### 496 **Author information**

497 The authors declare no competing interests.

498 **References**

- 499 1 Howe, L. & Wain, A. *Predicting the future*. V, 195 p. (Cambridge University Press,  
500 1993).
- 501 2 Bauer, P., Thorpe, A. & Brunet, G. The quiet revolution of numerical weather prediction.  
502 *Nature* **525**, 47, doi:10.1038/nature14956 (2015).
- 503 3 Hantson, S. *et al.* The status and challenge of global fire modelling. *Biogeosciences* **13**,  
504 3359-3375, doi:10.5194/bg-13-3359-2016 (2016).
- 505 4 Agapiou, A. Remote sensing heritage in a petabyte-scale: satellite data and heritage  
506 Earth Engine© applications. *Int. J Digit. Earth* **10**, 85-102 (2017).
- 507 5 Stockhause, M. & Lautenschlager, M. CMIP6 Data Citation of Evolving Data. *Data Sci. J*  
508 **16**, doi:10.5334/dsj-2017-030 (2017).
- 509 6 Lee, J., Weger, R. C., Sengupta, S. K. & Welch, R. M. A neural network approach to  
510 cloud classification. *IEEE T Geosci. Remote.* **28**, 846-855, doi:10.1109/36.58972 (1990).
- 511 7 Benediktsson, J. A., Swain, P. H. & Ersoy, O. K. Neural network approaches versus  
512 statistical methods in classification of multisource remote sensing data. *IEEE T Geosci. Remote.*  
513 **28**, 540-552, doi:10.1109/Tgrs.1990.572944 (1990).
- 514 8 Camps-Valls, G. & Bruzzone, L. *Kernel methods for remote sensing data analysis*. 434  
515 p. (John Wiley & Sons, 2009).
- 516 9 Gómez-Chova, L., Tuia, D., Moser, G. & Camps-Valls, G. Multimodal classification of  
517 remote sensing images: A review and future directions. *P IEEE* **103**, 1560-1584,  
518 doi:10.1109/Jproc.2015.2449668 (2015).
- 519 10 Camps-Valls, G., Tuia, D., Bruzzone, L. & Benediktsson, J. A. Advances in hyperspectral  
520 image classification: Earth monitoring with statistical learning methods. *IEEE Signal Proc.*  
521 *Mag.* **31**, 45-54 (2014). **[Comprehensive overview of machine learning for**  
522 **classification]**
- 523 11 Gislason, P. O., Benediktsson, J. A. & Sveinsson, J. R. Random Forests for land cover  
524 classification. *Pattern Recog. Lett.* **27**, 294-300, doi:10.1016/j.patrec.2005.08.011 (2006). **[One**  
525 **of the first machine learning papers for land cover classification, method now**  
526 **operationally used]**
- 527 12 Muhlbauer, A., McCoy, I. L. & Wood, R. Climatology of stratocumulus cloud  
528 morphologies: microphysical properties and radiative effects. *Atmos. Chem. Phys.* **14**, 6695-  
529 6716, doi:10.5194/acp-14-6695-2014 (2014).

- 530 13 Grimm, R., Behrens, T., Märker, M. & Elsenbeer, H. Soil organic carbon concentrations  
531 and stocks on Barro Colorado Island—Digital soil mapping using Random Forests analysis.  
532 *Geoderma* **146**, 102-113 (2008).
- 533 14 Hengl, T. et al. SoilGrids250m: Global gridded soil information based on machine  
534 learning. *PLoS One* **12**, e0169748, doi:10.1371/journal.pone.0169748 (2017). **[Machine**  
535 **learning used for operational global soil mapping]**
- 536 15 Townsend, P. A., Foster, J. R., Chastain, R. A. & Currie, W. S. Application of imaging  
537 spectroscopy to mapping canopy nitrogen in the forests of the central Appalachian Mountains  
538 using Hyperion and AVIRIS. *IEEE T Geosci. Remote.* **41**, 1347-1354,  
539 doi:10.1109/Tgrs.2003.813205 (2003).
- 540 16 Coops, N. C., Smith, M.-L., Martin, M. E. & Ollinger, S. V. Prediction of eucalypt foliage  
541 nitrogen content from satellite-derived hyperspectral data. *IEEE T Geosci. Remote.* **41**, 1338-  
542 1346, doi:10.1109/Tgrs.2003.813135 (2003).
- 543 17 Verrelst, J., Alonso, L., Camps-Valls, G., Delegido, J. & Moreno, J. Retrieval of  
544 vegetation biophysical parameters using Gaussian process techniques. *IEEE T Geosci.*  
545 *Remote.* **50**, 1832-1843, doi:10.1109/Tgrs.2011.2168962 (2012).
- 546 18 Papale, D. & Valentini, R. A new assessment of European forests carbon exchanges by  
547 eddy fluxes and artificial neural network spatialization. *Global Change Biol.* **9**, 525-535,  
548 doi:10.1046/j.1365-2486.2003.00609.x (2003).
- 549 19 Jung, M. et al. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat,  
550 and sensible heat derived from eddy covariance, satellite, and meteorological observations. *J*  
551 *Geophys. Res. - Biogeo.* **116**, G00j07, doi:10.1029/2010jg001566 (2011).
- 552 20 Tramontana, G. et al. Predicting carbon dioxide and energy fluxes across global  
553 FLUXNET sites with regression algorithms. *Biogeosciences* **13**, 4291 - 4313, doi:10.5194/bg-  
554 13-4291-2016 (2016).
- 555 21 Jung, M. et al. Recent decline in the global land evapotranspiration trend due to limited  
556 moisture supply. *Nature* **467**, 951 - 954, doi:10.1038/nature09396 (2010). **[First data-**  
557 **driven machine learning based spatio-temporal estimation of global water fluxes on**  
558 **land]**
- 559
- 560 22 Jung, M. et al. Compensatory water effects link yearly global land CO<sub>2</sub> sink changes to  
561 temperature. *Nature* **541**, 516 - 520, doi:10.1038/nature20780 (2017).
- 562 23 Bonan, G. B. et al. Improving canopy processes in the Community Land Model version 4  
563 (CLM4) using global flux fields empirically inferred from FLUXNET data. *J Geophys. Res. -*  
564 *Biogeo.* **116**, G02014, doi:10.1029/2010jg001593 (2011).

- 565 24 Anav, A. et al. Spatiotemporal patterns of terrestrial gross primary production: A review.  
566 *Rev. Geophys.* **53**, 785-818, doi:10.1002/2015rg000483 (2015).
- 567 25 Landschützer, P. et al. A neural network-based estimate of the seasonal to inter-annual  
568 variability of the Atlantic Ocean carbon sink. *Biogeosciences* **10**, 7793-7815, doi:10.5194/bg-10-  
569 7793-2013 (2013). **[First data-driven machine learning based carbon fluxes in the ocean]**
- 570 26 Kühnlein, M., Appelhans, T., Thies, B. & Nauss, T. Improving the accuracy of rainfall  
571 rates from optical satellite sensors with machine learning—A random forests-based approach  
572 applied to MSG SEVIRI. *Remote Sens. Environ.* **141**, 129-143 (2014).
- 573 27 Caldwell, P. M. et al. Statistical significance of climate sensitivity predictors obtained by  
574 data mining. *Geophys. Res. Lett.* **41**, 1803-1808, doi:10.1002/2014gl059205 (2014).
- 575 28 Reichstein, M. & Beer, C. Soil respiration across scales: the importance of a model-data  
576 integration framework for data interpretation. *J. Plant Nutr. Soil Sci.* **171**, 344 - 354,  
577 doi:10.1002/jpln.200700075 (2008).
- 578 29 Wright, S. Correlation and causation. *J. Agric. Res.* **20**, 557-585 (1921).
- 579 30 Guttman, N. B. Accepting the standardized precipitation index: a calculation algorithm. *J*  
580 *Am. Water. Resour. As.* **35**, 311-322, doi:10.1111/j.1752-1688.1999.tb03592.x (1999).
- 581 31 Vicente-Serrano, S. M., Beguería, S. & López-Moreno, J. I. A multiscalar drought index  
582 sensitive to global warming: the standardized precipitation evapotranspiration index. *J. Clim.* **23**,  
583 1696-1718, doi:10.1175/2009jcli2909.1 (2010).
- 584 32 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436-444,  
585 doi:10.1038/nature14539 (2015).
- 586 33 Lore, K. G., Stoecklein, D., Davies, M., Ganapathysubramanian, B. & Sarkar, S. in  
587 *JMLR: Workshop and Conference Proceeding / The 1st International Workshop "Feature*  
588 *Extraction: Modern Questions and Challenge"* Vol. 44 213-225 (Montreal, Canada, 2015).
- 589 34 Baldi, P., Sadowski, P. & Whiteson, D. Searching for exotic particles in high-energy  
590 physics with deep learning. *Nat. Commun.* **5**, 4308, doi:10.1038/ncomms5308 (2014).
- 591 35 Bhimji, W., Farrell, S. A., Kurth, T., Paganini, M. & Racah, E. Deep Neural Networks for  
592 Physics Analysis on low-level whole-detector data at the LHC. *arXiv.org e-Print archive*,  
593 arXiv:1711.03573 (2017).
- 594 36 Schutt, K. T., Arbabzadah, F., Chmiela, S., Muller, K. R. & Tkatchenko, A. Quantum-  
595 chemical insights from deep tensor neural networks. *Nat. Commun.* **8**, 13890,  
596 doi:10.1038/ncomms13890 (2017).

Reichstein et al., *Deep learning and process-understanding for data-driven Earth System science*

- 597 37 Alipanahi, B., DeLong, A., Weirauch, M. T. & Frey, B. J. Predicting the sequence  
598 specificities of DNA-and RNA-binding proteins by deep learning. *Nat. Biotechnol.* **33**, 831-838,  
599 doi:10.1038/nbt.3300 (2015).
- 600 38 Prabhat. *A Look at Deep Learning for Science*, O'Reilly Blog.,  
601 <https://www.oreilly.com/ideas/a-look-at-deep-learning-for-science> (2017).
- 602 39 Zhang, L. P., Zhang, L. F. & Du, B. Deep Learning for Remote Sensing Data A technical  
603 tutorial on the state of the art. *IEEE Geosc. Rem. Sen. M* **4**, 22-40,  
604 doi:10.1109/Mgrs.2016.2540798 (2016).
- 605 40 Ball, J. E., Anderson, D. T. & Chan, C. S. Comprehensive survey of deep learning in  
606 remote sensing: theories, tools, and challenges for the community. *J Appl. Remote Sens.* **11**,  
607 042609, doi:10.1117/1.Jrs.11.042609 (2017).
- 608 41 Racah, E. et al. ExtremeWeather: A large-scale climate dataset for semi-supervised  
609 detection, localization, and understanding of extreme weather events. *Adv. Neural. Inf. Process.*  
610 *Syst.*, 3405-3416 (2017).
- 611 42 Liu, Y. et al. in *ABDA'16 - International Conference on Advances in Big Data Analytics*  
612 81-88 (2016). **[First approach to automatically detect extreme weather without any**  
613 **prescribed thresholds, using deep learning]**
- 614 43 Zhao, W. Z. & Du, S. H. Learning multiscale and deep representations for classifying  
615 remotely sensed imagery. *ISPRS J Photogramm. Remote. Sens.* **113**, 155-165,  
616 doi:10.1016/j.isprsjrs.2016.01.004 (2016).
- 617 44 Mathieu, M., Couprie, C. & LeCun, Y. Deep multi-scale video prediction beyond mean  
618 square error. *arXiv.org e-Print archive*, arXiv:1511.05440 (2015).
- 619 45 Oh, J., Guo, X., Lee, H., Lewis, R. L. & Singh, S. Action-conditional video prediction  
620 using deep networks in atari games. *Adv. Neural. Inf. Process. Syst.*, 2863-2871 (2015).
- 621 46 Shi, X. et al. Convolutional LSTM Network: A Machine Learning Approach for  
622 Precipitation Nowcasting. *Adv. Neural. Inf. Process. Syst.* **28**, 802-810 (2015).
- 623 47 Deng, J. et al. in *2009 IEEE Conference on Computer Vision and Pattern Recognition*  
624 248-255 (IEEE, Miami, FL, 2009).
- 625 48 Friedlingstein, P. et al. Uncertainties in CMIP5 Climate Projections due to Carbon Cycle  
626 Feedbacks. *J. Clim.* **27**, 511-526, doi:10.1175/jcli-d-12-00579.1 (2014).
- 627 49 Montavon, G., Samek, W. & Müller, K.-R. Methods for interpreting and understanding  
628 deep neural networks. *Digit. Signal Process.* (2017).

- 629 50 Runge, J. et al. Identifying causal gateways and mediators in complex spatio-temporal  
630 systems. *Nat. Commun.* **6**, 8502, doi:10.1038/ncomms9502 (2015).
- 631 51 Chalupka, K., Bischoff, T., Perona, P. & Eberhardt, F. in *UAI'16 Proceedings of the*  
632 *Thirty-Second Conference on Uncertainty in Artificial Intelligence* 72-81 (AUAI Press Arlington,  
633 Jersey City, New Jersey, USA, 2016).
- 634 52 Ghahramani, Z. Probabilistic machine learning and artificial intelligence. *Nature* **521**,  
635 452, doi:10.1038/nature14541 (2015).
- 636 53 Goodfellow, I. J. et al. Generative Adversarial Nets. *Adv. Neural. Inf. Process. Syst.* **27**,  
637 2672-2680 (2014). **[Fundamental paper on a deep generative modelling approach,**  
638 **allowing one to model e.g. possible futures from data]**
- 639 54 Hansen, M. C. et al. High-resolution global maps of 21st-century forest cover change.  
640 *Science* **342**, 850-853, doi:10.1126/science.1244693 (2013).
- 641 55 Pekel, J.-F., Cottam, A., Gorelick, N. & Belward, A. S. High-resolution mapping of global  
642 surface water and its long-term changes. *Nature* **540**, 418-422, doi:10.1038/nature20584  
643 (2016).
- 644 56 Karpatne, A. et al. Theory-guided Data Science: A New Paradigm for Scientific  
645 Discovery from Data. *IEEE T Knowl. Data En.* **29**, 2318-2331, doi:10.1109/TKDE.2017.2720168  
646 (2017).
- 647 57 Camps-Valls, G. et al. Physics-aware Gaussian processes in remote sensing. *Appl. Soft*  
648 *Comput.* **68**, 69-82, doi:10.1016/j.asoc.2018.03.021 (2018).
- 649 58 Karpatne, A., Watkins, W., Read, J. & Kumar, V. Physics-guided Neural Networks  
650 (PGNN): An Application in Lake Temperature Modeling. *arXiv.org e-Print archive*,  
651 arXiv:1710.11431 (2017).
- 652 59 Luo, Y. Q. et al. A framework for benchmarking land models. *Biogeosciences* **9**, 3857 -  
653 3874, doi:10.5194/bg-9-3857-2012 (2012).
- 654 60 Eyring, V. et al. Towards improved and more routine Earth system model evaluation in  
655 CMIP. *Earth Syst. Dynam.* **7**, 813-830, doi:10.5194/esd-7-813-2016 (2016).
- 656 61 Klocke, D., Pincus, R. & Quaas, J. On constraining estimates of climate sensitivity with  
657 present-day observations through model weighting. *J. Clim.* **24**, 6092-6099,  
658 doi:10.1175/2011jcli4193.1 (2011).
- 659 62 Cox, P. M. et al. Sensitivity of tropical carbon to climate change constrained by carbon  
660 dioxide variability. *Nature* **494**, 341-344, doi:10.1038/nature11882 (2013).

- 661 63 Beck, H. E. et al. Global-scale regionalization of hydrologic model parameters. *Water*  
662 *Resour. Res.* **52**, 3599-3622 (2016).
- 663 64 Schirber, S., Klocke, D., Pincus, R., Quaas, J. & Anderson, J. L. Parameter estimation  
664 using data assimilation in an atmospheric general circulation model: From a perfect toward the  
665 real world. *J Adv. Model. Earth Systems* **5**, 58-70, doi:10.1029/2012ms000167 (2013).
- 666 65 Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G. & Yacalis, G. Could machine learning  
667 break the convection parameterization deadlock? *Geophys. Res. Lett.* **45**, 5742–5751,  
668 doi:10.1029/2018GL078202 (2018).
- 669 66 Becker, T., Stevens, B. & Hohenegger, C. Imprint of the convective parameterization  
670 and sea-surface temperature on large-scale convective self-aggregation. *J Adv. Model. Earth*  
671 *Systems* (2017).
- 672 67 Siongco, A. C., Hohenegger, C. & Stevens, B. Sensitivity of the summertime tropical  
673 Atlantic precipitation distribution to convective parameterization and model resolution in  
674 ECHAM6. *J Geophys. Res. - Atmos.* **122**, 2579-2594, doi:10.1002/2016jd026093 (2017).
- 675 68 de Bezenac, E., Pajot, A. & Gallinari, P. Deep Learning for Physical Processes:  
676 Incorporating Prior Scientific Knowledge. *arXiv.org e-Print archive*, arXiv:1711.07970 (2017).
- 677 69 Brenowitz, N. D. & Bretherton, C. S. Prognostic validation of a neural network unified  
678 physics parameterization. *Geophys. Res. Lett.* **45**, 6289-6298, doi:10.1029/2018gl078510  
679 (2018).
- 680 70 Willis, M. J. & von Stosch, M. Simultaneous parameter identification and discrimination  
681 of the nonparametric structure of hybrid semi-parametric models. *Comput. Chem. Eng.* **104**,  
682 366-376, doi:10.1016/j.compchemeng.2017.05.005 (2017).
- 683 71 McGovern, A. et al. Using artificial intelligence to improve real-time decision making for  
684 high-impact weather. *B Am. Meteorol. Soc.* **98**, 2073-2090, doi:10.1175/Bams-D-16-0123.1  
685 (2017).
- 686 72 Vandal, T. et al. in *Proceedings of the Twenty-Seventh International Joint Conference on*  
687 *Artificial Intelligence (IJCAI-18)* 5389-5393 (Stockholm, Sweden, 2018).
- 688 73 Verrelst, J. et al. Emulation of Leaf, Canopy and Atmosphere Radiative Transfer Models  
689 for Fast Global Sensitivity Analysis. *Remote Sens.* **8**, 673, doi:10.3390/rs8080673 (2016).
- 690 74 Chevallier, F., Chéruy, F., Scott, N. & Chédin, A. A neural network approach for a fast  
691 and accurate computation of a longwave radiative budget. *J Appl. Meteorol.* **37**, 1385-1397,  
692 doi:10.1175/1520-0450(1998)037<1385:Annafa>2.0.Co;2 (1998).
- 693 75 Castruccio, S. et al. Statistical emulation of climate model projections based on  
694 precomputed GCM runs. *J. Clim.* **27**, 1829-1844, doi:10.1175/Jcli-D-13-00099.1 (2014).

- 695 76 Fer, I. et al. Linking big models to big data: efficient ecosystem model calibration through  
696 Bayesian model emulation. *Biogeosci. Disc.* **2018**, 1-30, doi:10.5194/bg-2018-96 (2018).
- 697 77 Glahn, H. R. & Lowry, D. A. The use of model output statistics (MOS) in objective  
698 weather forecasting. *J Appl. Meteorol.* **11**, 1203-1211 (1972).
- 699 78 Wilks, D. S. Multivariate ensemble Model Output Statistics using empirical copulas. *Q J*  
700 *Roy. Meteor. Soc.* **141**, 945-952, doi:10.1002/qj.2414 (2015).
- 701 79 Tewari, A. et al. in *Proceedings of the IEEE Conference on Computer Vision and Pattern*  
702 *Recognition* 2549-2559 (2018).
- 703 80 Xie, Y., Franz, E., Chu, M. & Thuerey, N. tempoGAN: A Temporally Coherent,  
704 Volumetric GAN for Super-resolution Fluid Flow. *arXiv.org e-Print archive*, arXiv:1801.09710  
705 (2018).
- 706 81 Stewart, R. & Ermon, S. in *Proceedings of the Thirty-First AAAI Conference on Artificial*  
707 *Intelligence (AAAI-17)* 2576-2582 (San Francisco, California USA, 2017).
- 708 82 Gunning, D. *Explainable artificial intelligence (xai)*,  
709 [https://www.cc.gatech.edu/~alanwags/DLAI2016/\(Gunning\)%20IJCAI-16%20DLAI%20WS.pdf](https://www.cc.gatech.edu/~alanwags/DLAI2016/(Gunning)%20IJCAI-16%20DLAI%20WS.pdf)  
710 (2017).
- 711 83 Hu, Z., Ma, X., Liu, Z., Hovy, E. & Xing, E. in *Proceedings of the 54th Annual Meeting of*  
712 *the Association for Computational Linguistics* Vol. 1: Long Papers 2410–2420 (Association for  
713 Computational Linguistics, 2016).
- 714 84 Pearlmutter, B. A. & Siskind, J. M. Reverse-mode AD in a functional framework: Lambda  
715 the ultimate backpropagator. *ACM T Progr. Lang. Sys.* **30**, 7, doi:10.1145/1330017.1330018  
716 (2008).
- 717 85 Wang, F. & Rompf, T. in *ICLR 2018 Workshop* (2018).
- 718 86 Beer, C. et al. Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and  
719 Covariation with Climate. *Science* **329**, 834 - 838, doi:10.1126/science.1184984 (2010).
- 720 87 Forkel, M. et al. Enhanced seasonal CO<sub>2</sub> exchange caused by amplified plant  
721 productivity in northern ecosystems. *Science* **351**, 696-699, doi:10.1126/science.aac4971  
722 (2016).
- 723 88 Bellprat, O., Kotlarski, S., Lüthi, D. & Schär, C. Objective calibration of regional climate  
724 models. *J Geophys. Res. - Atmos.* **117**, D23115, doi:10.1029/2012jd018262 (2012).
- 725 89 Reichstein, M. et al. in *AGU Fall Meeting Abstracts* 2016AGUFM.B2044A..2007R  
726 (2016).



- 727 90 Rußwurm, M. & Körner, M. Multi-temporal land cover classification with long short-term  
728 memory neural networks. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XLII-1/W1**,  
729 551-558, doi:10.5194/isprs-archives-XLII-1-W1-551-2017 (2017). **[First Use of LSTM deep**  
730 **learning model for multi-temporal land-cover classification]**
- 731 91 Nash, J. E. & Sutcliffe, J. V. River flow forecasting through conceptual models part I—A  
732 discussion of principles. *J Hydrol.* **10**, 282-290 (1970).
- 733 92 Shi, X. et al. Deep Learning for Precipitation Nowcasting: A Benchmark and A New  
734 Model. *Adv. Neural. Inf. Process. Syst.* **30**, 5617-5627 (2017). **[First approach to data-driven**  
735 **modelling of near-term precipitation using a combination of deep-learning concepts, i.e.**  
736 **LSTMs and convolutional neural networks]**
- 737 93 Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional  
738 adversarial networks. *arXiv.org e-Print archive*, arXiv:1611.07004v07001 (2016). **[A geo-**  
739 **science related extension application of Goodfellow et al., where e.g. remote sensing**  
740 **images are transferred to thematic maps]**
- 741 94 Tompson, J., Schlachter, K., Sprechmann, P. & Perlin, K. in *Proceedings of the 34th*  
742 *International Conference on Machine Learning* Vol. 70 (eds Doina Precup & Yee Whye Teh)  
743 3424-3433 (PMLR, Proceedings of Machine Learning Research, 2017).
- 744 95 <https://nar.ucar.edu/2013/ral/short-term-explicit-prediction-step-program>
- 745 96 Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards real-time object detection  
746 with region proposal networks. *Adv. Neural. Inf. Process. Syst.*, 91-99 (2015).
- 747 97 Zaytar, M. A. & El Amrani, C. Sequence to sequence weather forecasting with long short  
748 term memory recurrent neural networks. *Int. J Comput. Appl.* **143** (2016).
- 749 98 Goodfellow, I., Bengio, Y. & Courville, A. *Deep learning*. xxii, 775 p. (MIT press, 2016).
- 750 99 May, R. M. Simple mathematical models with very complicated dynamics. *Nature* **261**,  
751 459-467, doi:10.1038/261459a0 (1976).
- 752 100 Siegelmann, H. T. & Sontag, E. D. On the computational power of neural nets. *J*  
753 *Comput. Syst. Sci.* **50**, 132-150, doi:10.1006/jcss.1995.1013 (1995).
- 754 101 Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735-  
755 1780 (1997).
- 756 102 Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **61**, 85-  
757 117, doi:10.1016/j.neunet.2014.09.003 (2015).
- 758

759 **Tables**

760

761 Table 1: Geoscientific tasks, conventional approaches, their limitations and potential of deep  
762 learning approaches

Analytical Task	Scientific Task	Conventional approaches	Limitations	Emergent or potential approaches
<b>Classification and anomaly detection</b>				
	Finding extreme weather patterns	Multivariate, threshold based detection	Heuristic approach, ad hoc criteria used	Supervised and Semi-supervised Convolutional Neural Networks <sup>41,42</sup>
	Land-use and change detection	Pixel-by-pixel spectral classification	No or only shallow spatial context used	Convolutional Neural Networks <sup>43</sup>
<b>Regression</b>				
	Predict fluxes from atmospheric conditions	Random forests Kernel methods Feedforward NNs	Memory and lag effects not considered	Recurrent neural networks, LSTMs <sup>89</sup>
	Predict vegetation properties from atmospheric conditions	Semi-empirical algorithms (temperature sums, water deficits)	Prescriptive in terms of functional forms and dynamic assumptions	Recurrent neural networks <sup>90</sup> , possibly with spatial context
	Predict river runoff in ungauged catchments	Process-models or statistical models with hand-designed topographic features <sup>91</sup>	Consideration of spatial context limited to hand-designed features	Combination of convolutional neural network with recurrent networks
<b>State Prediction</b>				
	Precipitation nowcasting	Physical modelling with data-assimilation	Computational limits due to resolution, data only used to update states	Convolutional-LSTM nets short-range spatial context <sup>92</sup>
	Downscaling and bias correcting forecasts	Dynamic modelling and statistical approaches	Computational limits; subjective feature selection	Convolutional nets <sup>72</sup> , cGANs <sup>53,93</sup>
	Seasonal forecasts	Physical modelling with initial conditions from data	Fully dependent on physical model, current skill relatively weak	Convolutional-LSTM nets with long-range spatial context
	Transport modelling	Physical modelling of transport	Fully dependent on physical model, computational limits	Hybrid physical-convolutional network models <sup>94, 68</sup>

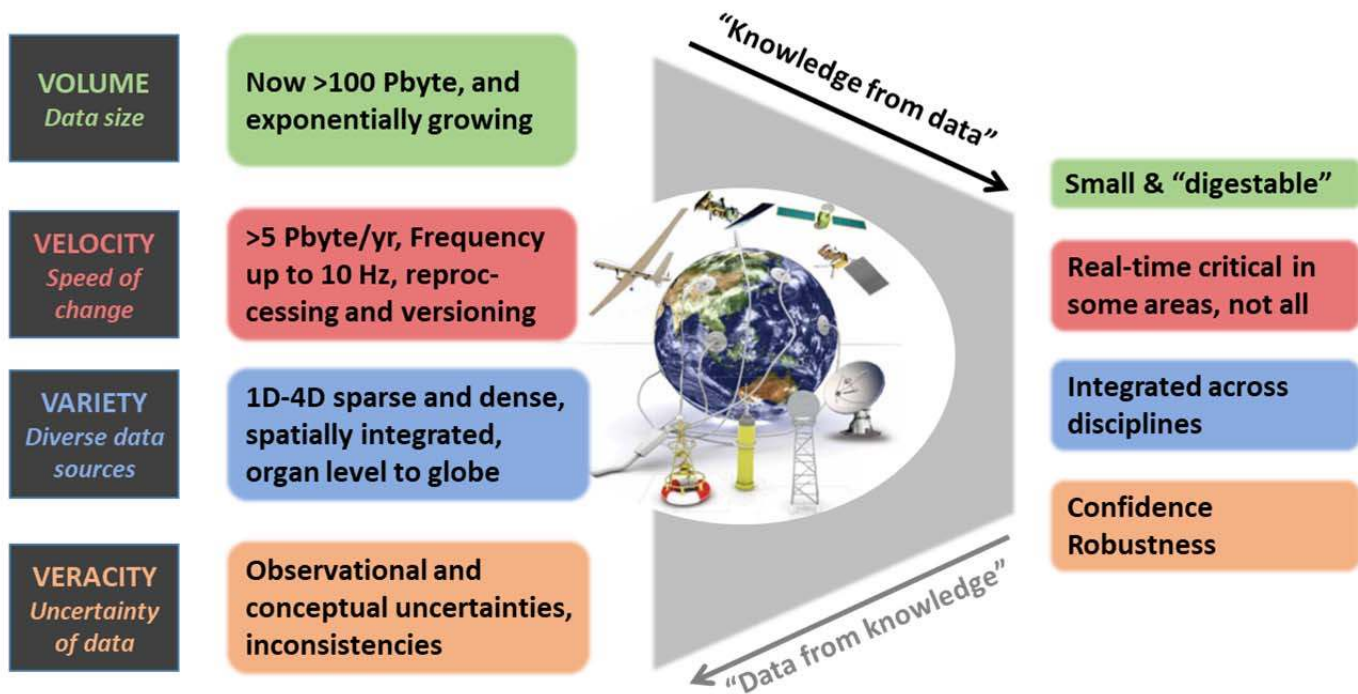
763

764  
765  
766  
767  
768  
769

**Figures with captions**

**Big data: Observed and simulated**

**Patterns & knowledge**

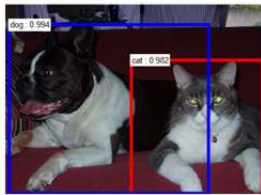


770  
771  
772  
773  
774  
775

**Figure 1: Big data challenges in the geoscientific context** (Earth picture from <https://nosc.noaa.gov/tpio/images/ObsSys.jpg>)

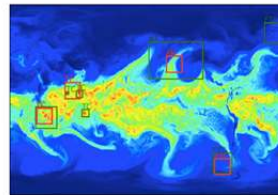
## Machine learning tasks

### Object classification and localization

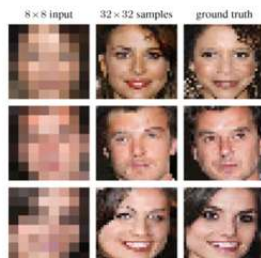


## Earth science tasks

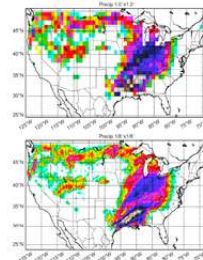
### Pattern classification



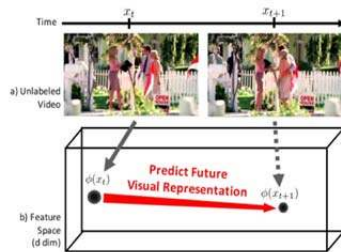
### Super-resolution and fusion



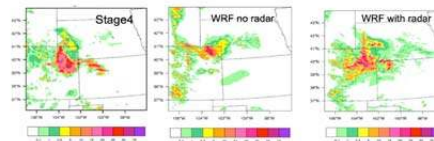
### Statistical downscaling and blending



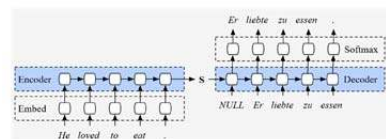
### Video prediction



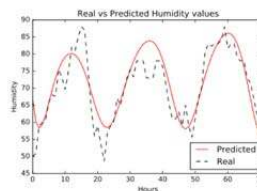
### Short-term forecasting



### Language translation



### Dynamic time series modeling



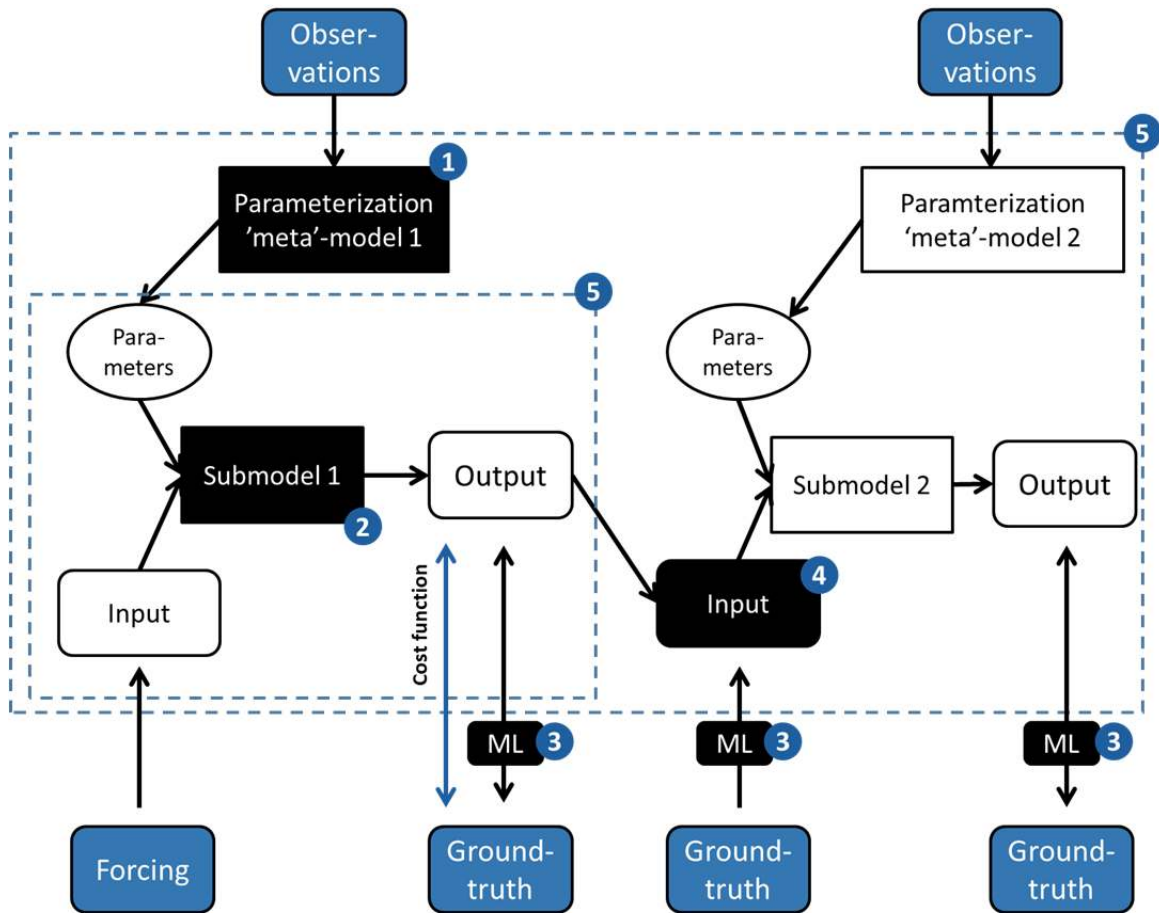
776

777 **Figure 2: Typical deep learning applications (left) and the geo-scientific problems they**  
 778 **apply to (right).** From top to bottom: 1) classification of extreme weather patterns using a  
 779 unified convolutional neural network on climate simulation data<sup>42, 41</sup>, 2) statistical downscaling of  
 780 climate model output<sup>72</sup>, 3) short-term forecasting of climate variables<sup>95</sup>, and 4) modelling of  
 781 dynamic time-series.<sup>96, 97</sup> Image sources: [https://smerity.com/articles/2016/google\\_nmt\\_arch.html](https://smerity.com/articles/2016/google_nmt_arch.html);  
 782 <https://arxiv.org/abs/1612.02095>;

783

784

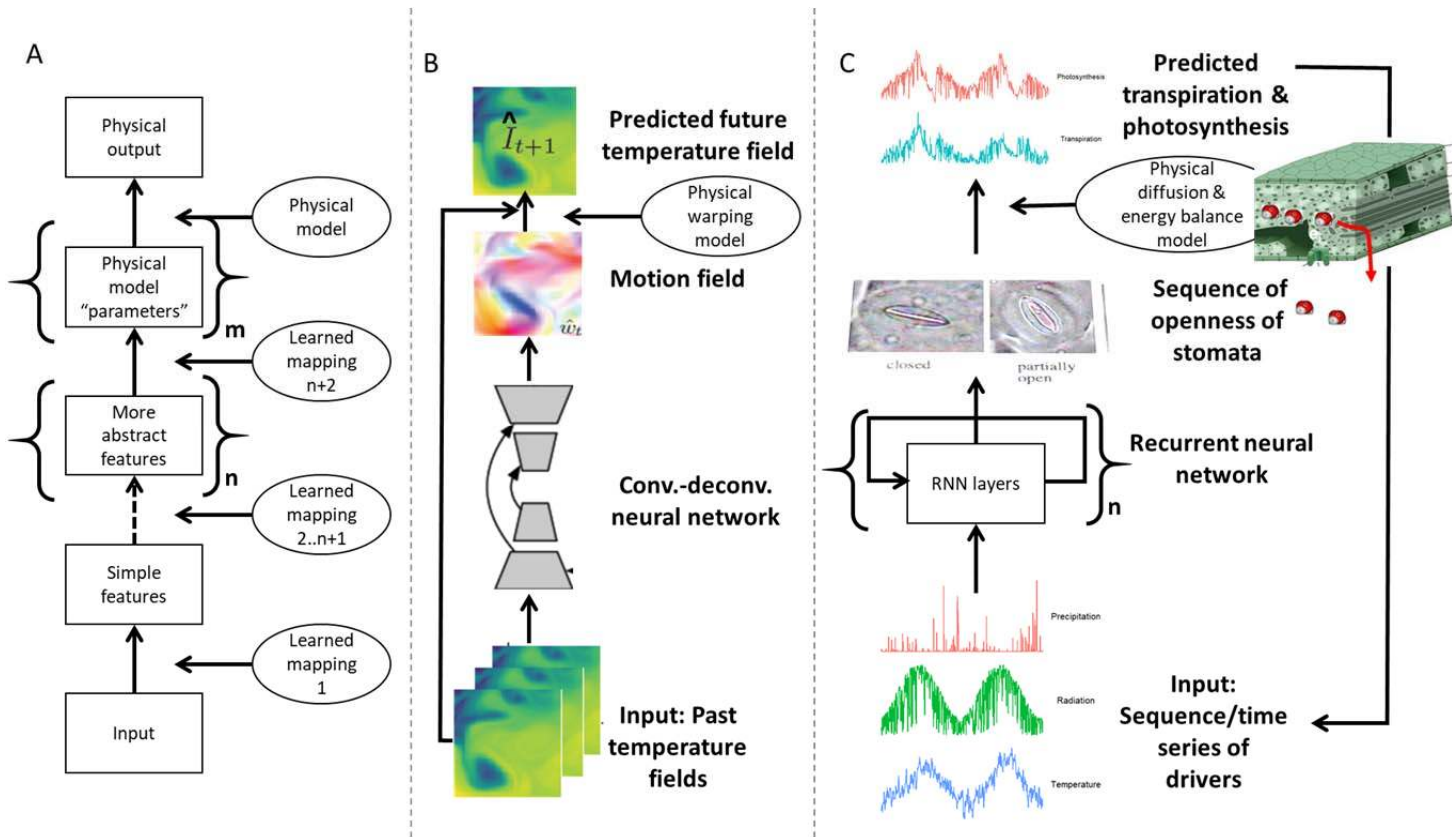
785



786

787 **Figure 3: Linkages between physical models and machine learning:** Depicted here is an  
 788 abstraction of a part of a physical system, e.g. a climate model. The model consists of  
 789 submodels which each have parameters, and forcing variables as inputs, and produce output,  
 790 which can be input (forcing) to another sub-model. Data-driven learning approaches can be  
 791 helpful in various instances, cf. the black-boxes and numbers. More detail in the text. ML =  
 792 Machine Learning

793



794

795 **Figure 4: Interpretation of hybrid modelling (circle 2 in Figure 3) as deepening and**  
 796 **“physicsizing” a deep learning architecture** by adding one or several ( $m$ ) physical layers  
 797 after the multilayer neural network (A). (B) and (C) are concrete examples, where (B) is from de  
 798 Bezenac et al.<sup>68</sup>, where a motion field is learned with a convolutional-deconvolutional neural  
 799 network, and the motion field further processed with a physical model. (C) models a biological  
 800 regulation process (opening of the stomatal “valves” controlling water vapor flux from the  
 801 leaves) with a recurrent neural network and processes this further with a physical diffusion  
 802 model to estimate transpiration, which in turn influences some of the drivers, e.g. soil moisture.  
 803 Basic scheme (A) modified after Goodfellow et al.<sup>98</sup>.

804

805 **Supplementary material and glossary**806 **Efficient modelling a dynamic non-linear system with recurrent neural networks**

807 Aforementioned state-of-the-art examples of mapping sequences of driving variables (e.g.  
 808 meteorological conditions) onto target variables such as CO<sub>2</sub> fluxes from ocean or land have considered  
 809 instantaneous mapping without representation of state dynamics. Dynamic effects have either been  
 810 considered by directly using observed states as predictors (e.g. vegetation state represented by  
 811 reflectance) or by introducing hand-designed features. The general problem is depicted in the figure  
 812 below, where the input acts on an unknown, unobservable system state, while the observable is both  
 813 influenced by the past state and the current input. It is not a problem of forecasting a time series a few  
 814 steps ahead, because the whole output sequence has to be predicted by the model.

815 As an example, in the synthetic dynamic system below (one realization in Figure Box 1) we have three  
 816 forcing variables  $x_1, x_2, x_3$  where two of them influence one (unobserved) state  $r$  according to

$$817 \quad r_{t+1} = f(x_{1,t}, x_{2,t}, r_t), \text{ with}$$

$$818 \quad f(x_{1,t}, x_{2,t}, r_t) = \tau \cdot x_{1,t} \cdot x_{2,t} \cdot e^{x_{1,t}} + (1 - \tau) \cdot r_t,$$

819  $\tau$  being a parameter determining the inertia of the dynamics of  $r$ , here set to 0.05. A target state  $y$  to be  
 820 predicted evolves as a logistic map well known from ecology and chaos theory<sup>99</sup>:

$$821 \quad y_{t+1} = \tilde{r}_t \cdot y_t \cdot (1 - y_t),$$

822 where (contrary to the standard logistic map) the parameter  $\tilde{r}_t$  is not fixed but dynamic and dependent  
 823 on  $r$  as

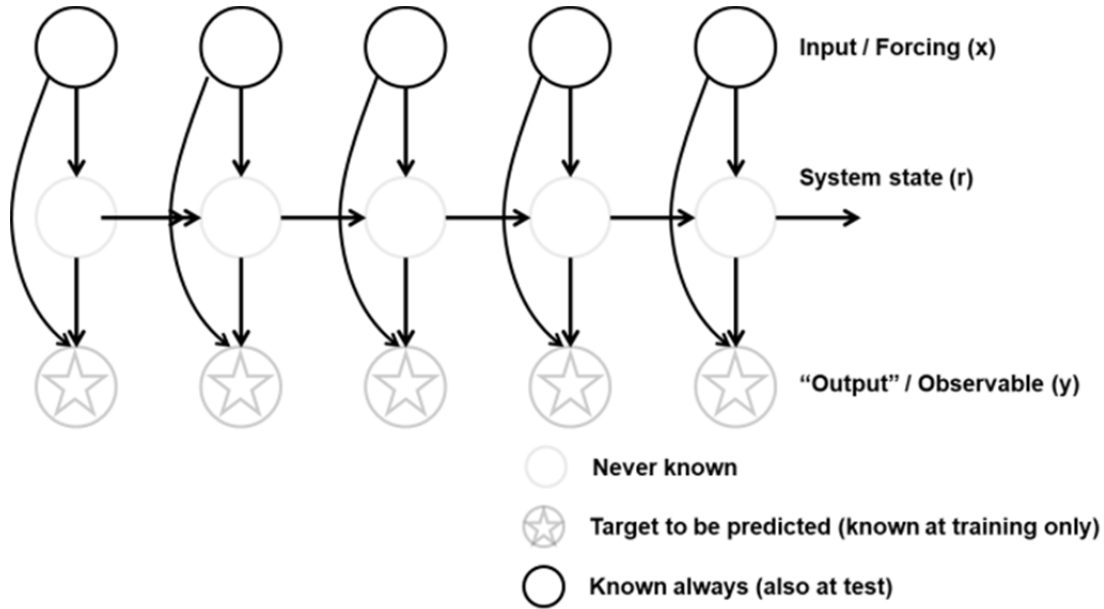
$$824 \quad \tilde{r}_t = g(r_t + x_{3,t}),$$

825 where  $g$  simply scales  $\tilde{r}_t$  onto the interval [2.5, 4] which implies dynamics varying with time between  
 826 dampened oscillations, limit cycles and chaos. In the synthetic example 500 realizations of  $x_1$  and  $x_2$  as  
 827 Gaussian i.i.d. variables are generated, while  $x_3$  is always a seasonal variable as in the Figure below.  
 828 Obviously  $x_1$  and  $x_2$  are mimicking a stochastic forcing, whereas  $x_3$  represents a deterministic forcing  
 829 (e.g. solar radiation varying diurnally and seasonally).

830 The lower panel shows the performance of different approaches to model the  $y_t$  sequence given the  
 831 sequences of  $x_1 \dots x_3$ . With a feed-forward ANN or random forests it is hard to model the sequence  $y_t$ ,  
 832 even with including intuitive features which represent lagged or memory effects, such as lagged or  
 833 cumulated  $x$  variables over the last 25 time steps. On the contrary, being turing-complete<sup>100</sup> a recurrent  
 834 NN has the potential to describe any dynamic system, and the challenge is the parameter estimation or  
 835 training. In the specific case a simple LSTM<sup>101</sup> with 8 cells was trained on 80% of the realizations and the  
 836 results are shown here for the test set. Certainly, other modelling approaches such as dynamic Bayesian

837 approaches (e.g. hidden Markov models) exist as well for state estimation, and the relation to recurrent  
838 neural networks and deep learning is under research<sup>102</sup>.

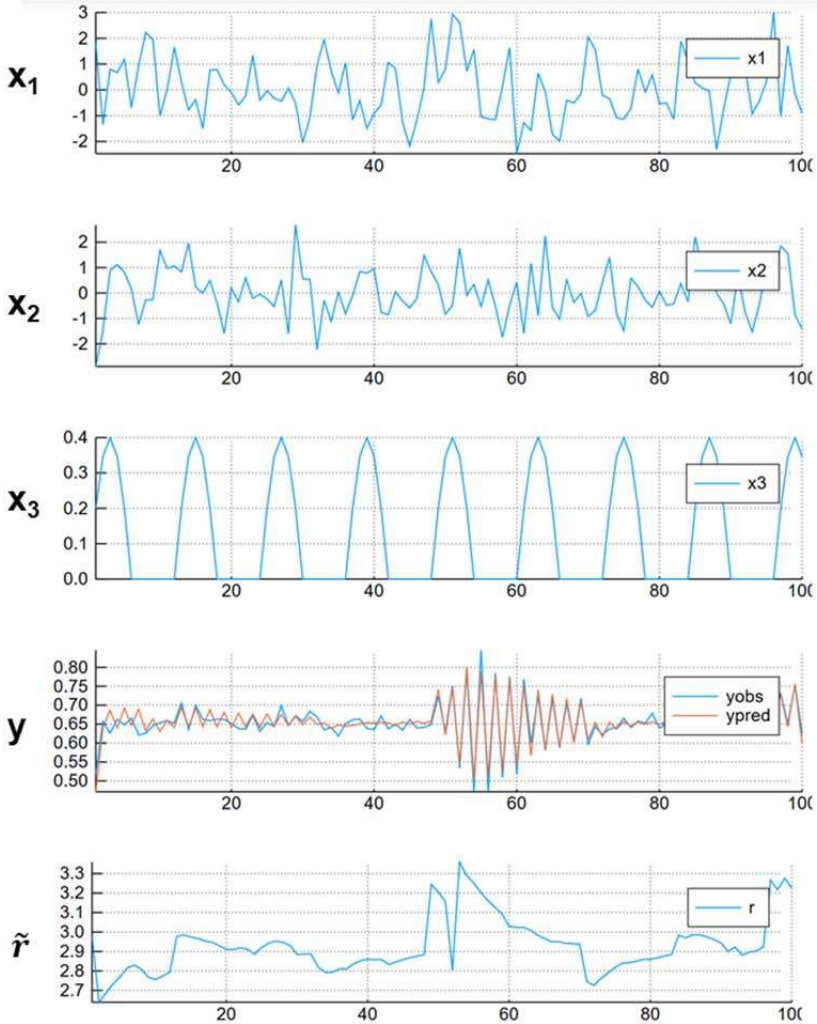
839



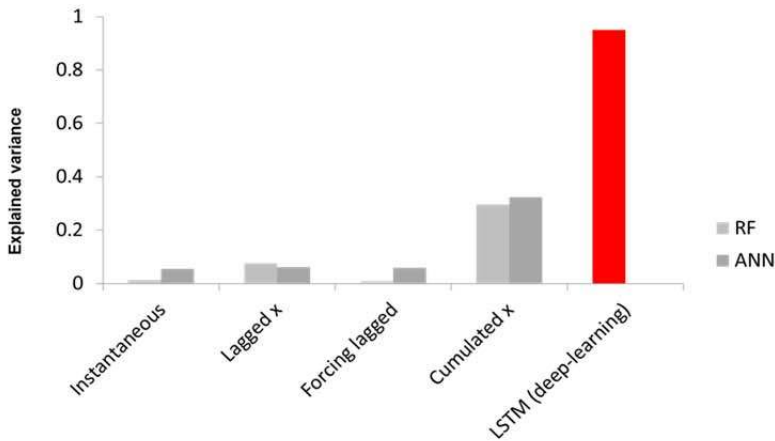
840

841 Supplementary Figure 1: Concept of modelling a dynamic system, i.e. mapping an input sequence to an  
842 output sequence, where a (hidden) dynamic state is involved.





843

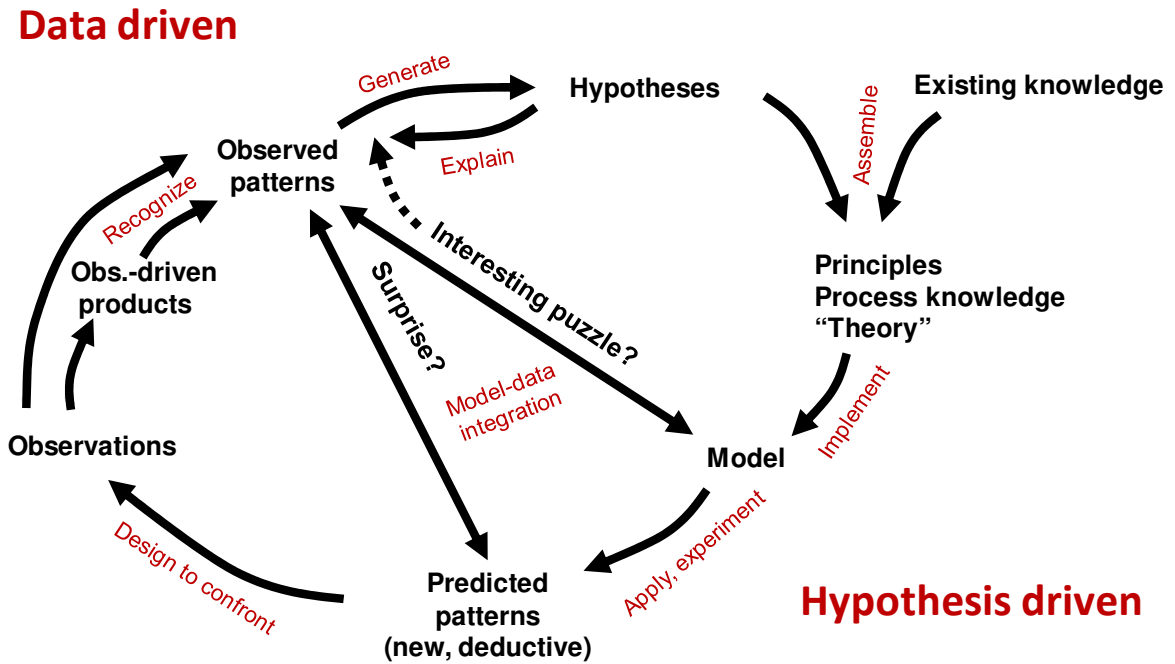


844

845 Supplementary Figure 2: Data-driven modelling of a synthetic geoscientific time-series depicted in (a) with  
 846 dynamic effects. Shown are predictor variables  $x_1$ ,  $x_2$ ,  $x_3$ , the resulting time-series of the system state  
 847 (“observed” and modelled with an LSTM), and the parameter  $\tilde{r}_t$  of the logistic map. (b) While classical  
 848 approaches including typical feature design fail to explain the dynamics (grey bars, RF = Random Forest,

849 ANN = feedforward ANN), a deep learning approach, long-short-term-memory neural network (LSTM) is  
850 able to explain almost all variance (red), without designing any features.

851



852

853 Supplementary Figure 3: Cycle of hypothesis generation and testing in complex systems  
854 involving process-based models and extraction of patterns from observations. Such patterns are  
855 only a surprise, and constitute an puzzle, if state-of-the-art theory and models do not predict  
856 them. Machine learning allows to extract hidden and complex patterns, which should be  
857 confronted with modelled patterns.

858

**Glossary**

<b>Term</b>	<b>Explanation</b>
Artificial Intelligence, Machine Learning & Deep Learning	<p>Artificial intelligence (AI) is the capacity of an algorithm for assimilating information to perform tasks that are characteristic of human intelligence, such as recognizing objects and sounds, contextualizing language, learning from the environment, and problem solving.</p> <p>Machine learning (ML) is a field of statistical research for training computational algorithms that split, sort, transform a set of data in order to maximize the ability to classify, predict, cluster or discover new patterns in target datasets.</p> <p>Deep learning refers to ML algorithms that construct hierarchical architectures of increasing sophistication. Artificial neural networks with many layers are examples of deep learning algorithms.</p>
Bayesian inference	Bayesian inference is a field in statistics and machine learning that develop methods for data analysis based on updating the probability for an hypothesis based on observational evidence. The framework is mostly concerned about treating uncertainty, encoding prior beliefs and estimating error propagation when dealing with data and models.
Causal inference	Causal inference links events, processes or properties in a system via a cause-effect connection. Recent observational causal inference try to discover causal relations from data.
Convolution	Convolution is one of the most important operations in signal and image processing, and it can operate in 1D (e.g. speech), 2D (e.g. images) or 3D (e.g. video) objects. A convolutional filter is essentially a weight vector/matrix/cube that operates in a sliding window approach on the data. Depending on the kernel structure, the operation enhances some features of the data, such as edges, trends, or flat regions. The operation is embedded in convolutional neural networks at the neuron level, which extracts useful features from the previous layers.
Differentiable programming	Differentiable programming refers to a programming paradigm to generate code that is automatically differentiated, such that its parameters can be seamlessly optimized. It generalizes current deep learning frameworks to arbitrary programs which may include the hybrid modelling approaches we discuss in section 5.
Feedforward vs Recurrent networks	An artificial neural network (ANN) is a computational algorithm that simulates how signals are transferred between a network of neurons, via synapses. In an ANN, information is transferred only in the forward direction while in a recurrent ANN the information can cycle/loop between the different nodes, creating complex dynamics, like memory, as seen in data.
Generative Adversarial Networks	Family of unsupervised ML methods widely used to generate

(GAN)	realistic samples from an unknown probability density function. GANs are formed by a neural network that generates plausible examples that try to fool a discriminator network that should discern real from fake examples.
Memory effects	Metaphoric term, meaning that the current behavior of a system cannot be explained without considering the effect of past states or forcing variables.
Nowcasting & Forecasting	To forecast a certain variable refers to establish a prediction of its value in the future, from days to centuries. Nowcasting refers to making that prediction in a very near future (e.g. predicting if it is going to rain in a couple of hours).
Probabilistic programming	Probabilistic programming is an approach to define probabilistic models with a unified high-level programming language. Statistical inference is automatically achieved by built-in inference machines, freeing the developer from the difficulties of high-performance probabilistic inference.
Radiative transfer models (RTMs)	Mathematical models that describe how radiation at different wavelengths (e.g. visible light) propagates through different mediums (e.g. atmosphere, vegetation canopy) by simulating absorption, emission, transmission and scattering processes.
Remote sensing	Remote sensing deals with measuring the radiance at different wavelengths reflected or emitted from an object or surface. Remote sensing uses satellite or airborne sensors to detect and classify objects as well as to estimate geo-scientific variables of interest (temperature, salinity or carbon dioxide), based on propagated reflectance signals (e.g. electromagnetic radiation).
Supervised & Unsupervised learning	In supervised learning an algorithm learns the input-to-output relationship by being provided both the inputs and the respective outputs, e.g. a set of photos (inputs) and a set of corresponding labels (outputs). In unsupervised learning the algorithms do not have access to the labels, so the goal is to infer the underlying structure of the data (e.g. the algorithm automatically separates pictures with different statistical or even semantic properties, e.g. images of cats and dogs).
Teleconnections	Teleconnections refer to climate anomalies related to each other at large distances (typically thousands of kilometers). Quantifying teleconnection patterns allows predicting key patterns on Earth, which are distant in space and time: e.g. predicting El Niño enables prediction of North American rainfall, snowfall, droughts or temperature patterns with a few weeks to months lead time.

860  
861  
862  
863

See <https://developers.google.com/machine-learning/glossary/> and <http://www.wildml.com/deep-learning-glossary/> for more complete glossaries.

864

865