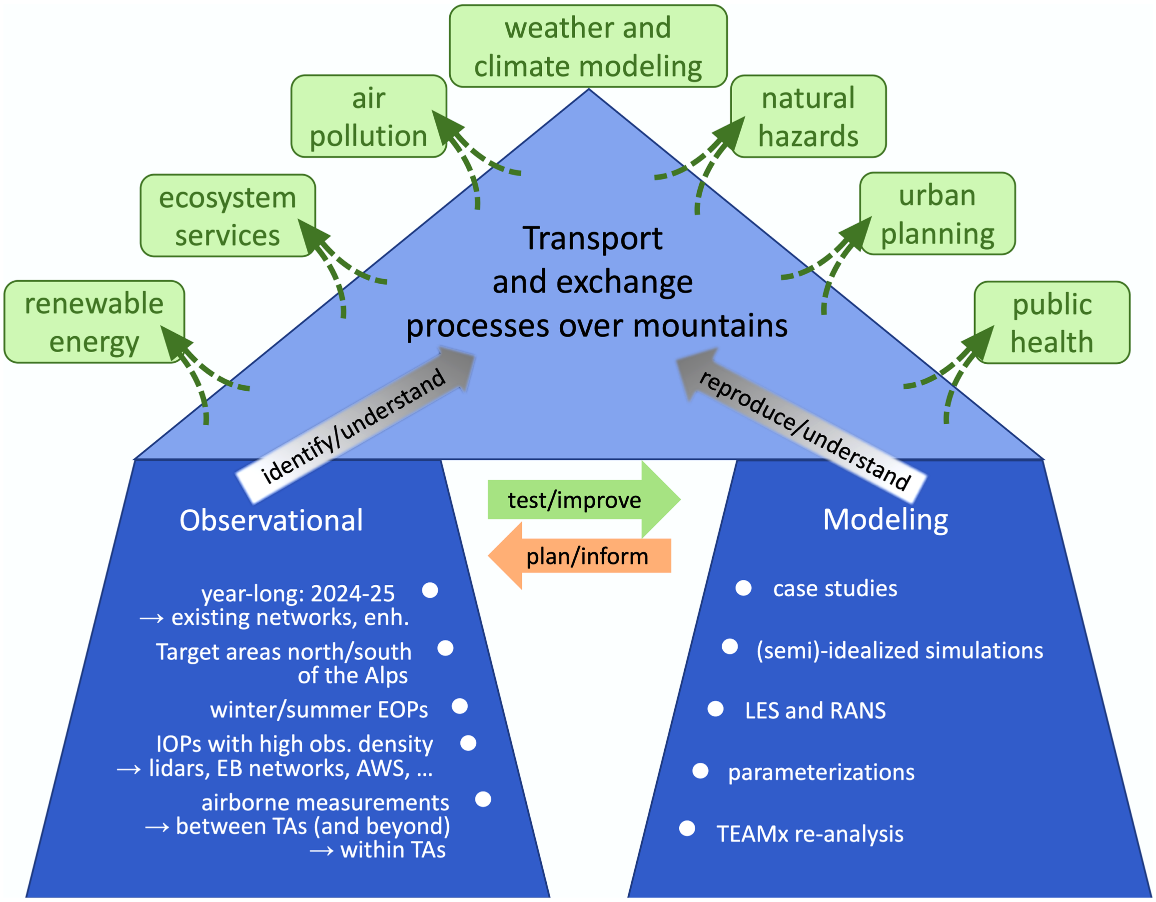

Fig. 2. Schematic representation of observational and numerical modeling activities planned within TEAMx. The TOC dataset is used to identify, characterize, and parameterize the relevant physical processes and to test and improve numerical modeling (different model types and parameterizations). Numerical modeling serves to optimize the experimental arrangements in the field, to specify the type of IOPs to be executed on a given day, and to improve data assimilation schemes and procedures in complex terrain, but it can also be used to identify the relative importance of different processes, assess uncertainties, and hence better understand the physical processes. "Case studies" refers to simulations with real terrain and atmospheric states close to reality (and thus includes short-term weather forecast simulations as well as climate time scales). LES denotes large-eddy simulation, RANS is Reynolds-averaged Navier–Stokes modeling (i.e., the typical framework for mesoscale atmospheric models), EOP is extended observation period, IOP is intensive observation period, TA is target area, EB is energy balance, AWS is automatic weather station, and "enh" is enhanced. Green boxes refer to typical areas of application for local atmospheric information.

Hand-in-hand with the TOC, a variety of numerical modeling experiments will be carried out across a range of scales and for different applications (Fig. 2). Central to TEAMx efforts is assessing the accuracy of model output with respect to understanding of exchange processes—it has become custom to call this "right for the right reason." To provide useful data for weather and climate services, we must have confidence that model output is sufficiently accurate for its purpose. For example, only with an adequate treatment and understanding of the surface energy balance in a complex valley can we correctly assess local evapotranspiration, the interaction of the locally driven slope flow with the mesoscale valley circulation, the interaction of the synoptic flow with both slope and valley flows, and hence the efficiency of the valley system's exchange of water vapor with the free troposphere. This improved representation of local evapotranspiration may in turn modify the abundance of water availability for precipitation processes at the mountain scale or even downstream. The key here is that the "right reason" (in this case, the correct surface energy balance, proper slope flow characteristics, etc.) not only yields the correct impact of the valley system on the atmospheric conditions aloft, but also allows for an adequate local forecast (or diagnostic), and even enables scientifically sound sensitivity studies, such as investigating the impact of forested versus non-forested slopes or urbanization in a valley.

Modeling of mountain atmospheres at high resolution can be done in at least two different ways: real-terrain (real atmosphere) simulations as in NWP or climate scenarios and idealized-orography simulations. The latter approach has been used widely to investigate idealized features of mountain landscapes, such as an infinite slope [e.g., the famous Prandtl (1942) model], the flow in a valley between two parallel straight ridges (e.g., Schmidli 2013), or the development of convective precipitation therein (e.g., Panosetti et al. 2016). In TEAMx both these approaches will be pursued in a coordinated manner.

Weather and climate services in the mountains

Point forecasts (and diagnostics) are an essential ingredient of weather and climate services. Many applied simulation tools in Earth system modeling (such as hydrological runoff modeling, pollutant dispersion modeling, sustainable energy potential assessment modeling, agricultural modeling) share a heavy dependence on meteorological input data and were originally developed, trained, and validated using meteorological station data. With the advance of computing power—and hence the increased resolution of meteorological models—these tools are being increasingly forced with the output of atmospheric models which dramatically extends their coverage. NWP model output (which is available every day, under any circumstance) is no longer exclusively employed for point weather forecasts, but also for a myriad of Earth system modeling applications used to make real-world decisions. Thus, the NWP output must be correct, or at least of sufficient accuracy, at every single grid point of application. The same is true if we use those Earth system models in conjunction with climate scenario simulations (which are also approaching convection-permitting resolution; e.g., Berthou et al. 2020; Ban et al. 2021), for example, to estimate the occurrence of high-impact weather (WMO 2017) under future climate scenarios, the availability of sustainable energy resources by the end of the century, or the possible impact of changes in agricultural practice.

Notably, a large portion of these Earth system modeling applications consider processes *that specifically occur in mountainous terrain*, despite difficulties in generating reliable point forecasts in the mountains (see below). This is true for

- surface runoff processes relevant for flash flood forecasts, hydropower planning, and operations;
- many hydrological processes involving snow and ice, such as planning for snow availability for tourism, securing drinking water availability downstream of major mountain ranges, avalanche forecasts, and road safety in icy conditions;
- processes affecting the siting of wind energy plants; and
- numerous air quality issues related to terrain, such as cold air pools and corresponding pollutant accumulation, and pollutant dispersion.

Overall objectives of TEAMx

Taking all of the above issues into account, TEAMx has four major objectives (Serafin et al. 2020), namely,

- to improve qualitative and quantitative understanding of transport and exchange processes both between the surface and atmosphere and at multiple scales within the atmosphere,
- to provide a unique observational dataset which can be used to investigate the wide range of transport and exchange processes in mountainous terrain and their spatiotemporal variability,
- to evaluate and improve the performance of weather and climate models over mountainous terrain, and
- to reduce errors in impact models by transferring the knowledge gained to weather and climate service providers.

These broad objectives are based on the implicit assumption that our limited understanding of the exchange processes over mountains leads to *reduced forecast quality in the mountains*. Correspondingly, the third and fourth objectives could be reformulated as the goal to make weather and climate simulations over mountainous terrain at least as accurate as over flat terrain and that weather and climate services will no longer be limited by errors in weather and climate information. Issues with the accuracy of simulations over mountains may not, necessarily, impact everyday weather forecasts (if appropriately communicated) or even climate scenarios (if appropriately averaged, bias corrected, and communicated). They will, however, have an important impact on the applications in Earth system modeling. If the deviations are systematic, they will have a multiplying effect on possible enhancement/reduction in exchange efficiency over mountains and hence possibly a further modifying impact on larger-scale atmospheric states (see the orographic drag example in sidebar “Is exchange over mountains relevant?”). All the above conjectures will be critically examined through the research efforts of TEAMx—which hopefully will result in a more reliable estimation of the exchange of energy, mass, and momentum over mountains. Although the amount of relevant literature on the quality of weather and climate information over mountains and the underlying reasons remains limited, we show below that the existing evidence generally supports these ideas.

Is the quality of weather forecasts (climate diagnostics) worse in the mountains?

Although one might expect that mountainous regions may exhibit higher predictability due to the presence of a stationary obstacle in the flow (e.g., Anthes et al. 1985), in reality model errors likely dominate this effect to give lower forecast skill for many variables. It is clear, however, that the task of producing accurate forecasts in mountainous terrain is at least more challenging than over HHF terrain when it comes to properly representing exchange processes. Flat terrain can be inhomogeneous (e.g., a coastal area) or complex in terms of surface cover and form (e.g., a metropolitan area), which will impact the exchange. However, if substantial orography is involved we face additional numerical modeling issues (Chow et al. 2019) such as vertical coordinate definition, inconsistency of some physical parameterizations, a higher degree of nonlinearity and hence chaos, and, for very high resolution, the possibility of numerical instabilities due to too steep slopes.

But do these challenges impact forecast quality? Relatively little can be found in the scientific literature concerning systematic differences between forecast quality at mountain sites and in flat terrain. While virtually all operational national meteorological and hydrological services verify their weather forecasts, there are no systematic verifications for “mountainous” versus “flat terrain” sites even in countries containing significant mountains. Figure 3 shows a very straightforward comparison of this kind for the forecast model of ECMWF (IFS) using 3 years of data and a very simple definition of mountainous terrain [station height $>$ 1000 m above mean sea level (MSL)] and flat terrain (station height $<$ 500 m MSL). Clearly, this is only a very crude distinction between flat and mountainous terrain and hence a very first attempt, but still, the example shows that the flat terrain sites score substantially better than the mountainous sites. As this is only one example, it does not mean that the same results will be obtained for other models, or mountain ranges, or time periods. It is, however, a first step for the more systematic evaluation that will be performed within TEAMx.

A similar story applies to the assessment of the reliability of climate information over mountains: no systematic evaluation of "mountain versus non-mountain sites" can be found in

Is exchange over mountains relevant?

Four examples at different spatial scales demonstrate the relevance of surface–atmosphere exchange in mountain areas for the reliability of weather forecasts, climate modeling, and impact modeling.

Large-scale example: Momentum exchange

Orography slows down the large-scale flow through a number of processes such as orographic flow blocking, excitation of (non) hydrostatic gravity waves, or flow deformation (form drag) at turbulence scales—the impact of which is usually summarized as drag exerted on the flow. Numerical models used for weather or climate simulations cannot typically resolve mountains with horizontal scales less than a few tens to hundreds of kilometers and, thus, the orographic drag needs to be parameterized. Since their introduction in the 1980s, orographic drag parameterizations have played a major role for the accuracy of the models' momentum budgets (and hence wind forecast accuracy).

In an IUGG centennial event (100 Years of Atmospheric Research) in 2019, the Head of Research of ECMWF showed that orographic drag ("GWD O. Roughness" in Fig. SB1) was among the top 10 most important improvements in the history of numerical weather prediction—of similar importance as data assimilation procedures, satellite data, etc. The major impact can be seen on the Northern Hemisphere (full red line) where most major mountain ranges are found.

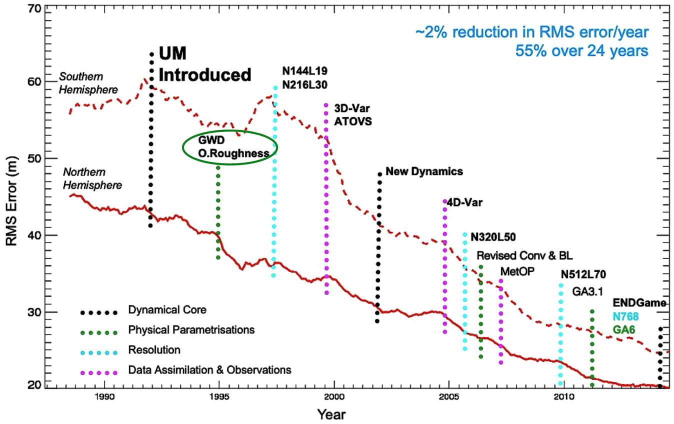

Fig. SB1. RMS error (vs analyses) of 500-hPa geopotential height, day 3 forecast by the Unified Model (UM) of the Met Office. Red dashed line for the Southern Hemisphere, red full line for the Northern Hemisphere. "GWD O. Roughness" (in the green ellipse) refers to the introduction of a gravity wave drag/orographic roughness parameterization. Figure courtesy Sean Milton from the Met Office, whose original work (Milton and Wilson 1996) is the basis of the shown impact.

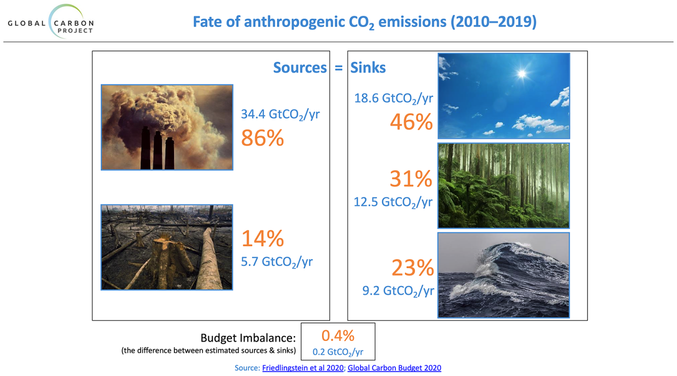

Fig. SB2. Summary of the fate of the anthropogenic CO_2 emissions (2010–19) from the Global Carbon Project (based on Friedlingstein et al. 2020). The sinks due to land surface exchange (middle panel to the right, 31%) are considered to be most uncertain. Source: www.globalcarbonproject.org/carbonbudget/20/presentation.htm, last accessed 1 Nov 2021.

Mesoscale example: Carbon budget

Due to its importance for many aspects of climate change, the global carbon budget is assessed every year by the internationally backed Global Carbon Project using the most up-to-date knowledge, data, and modeling tools to estimate anthropogenic emissions of greenhouse gases and their fate in the atmosphere and terrestrial/oceanic sinks. Every year, terrestrial processes—whether sources due to land-use change or sinks due to photosynthesis or other vegetative processes—are considered most uncertain (Fig. SB2). In fact, before 2016, the model-related uncertainties were considered so large that the terrestrial carbon sink was estimated as the residual of the carbon budget. More recently, the terrestrial sink is estimated based on atmospheric inversions with a resolution of atmospheric forcing of order $>$100 km. Note that the CO_2 flux data used are mostly not from orographically influenced sites (Rotach et al. 2014). Thus, at least part of the uncertainty may be attributed to the missing orography-related processes (Rotach et al. 2014), rendering the global budget compromised by uncertainties in mesoscale exchange processes.