A. DESCRIPTION OF PROPOSED RESEARCH

Title: Towards a high-fidelity Integrated Forecasting System via ground-breaking and ambitious data-assimilation of the dynamic soil-vegetation hydraulic continuum

1. Rationale: High-fidelity global-scale monitoring and modelling of the spatio-temporal variability of **land surface processes** and state variables, and their interactions and feedbacks with the atmosphere, is of crucial importance for reliable Numerical Weather Prediction (NWP) and climate modelling. In this context, **key land surface state variables** (LSSV) are the land surface temperature (LST, a complex mix of vegetation and soil surface temperatures), the near surface and rootzone soil moisture content, soil temperature (T_{soil}), snow albedo and snow density. LSSV play a pivotal role in the energy-, water and carbon balance, as well as in the strength of Land-Atmosphere (L-A) coupling, and how this is affected by soil moisture- (Song et al., 2019) and enthalpy (Hu and Feng, 2004) **'memories'**. The LSSV, and the **land surface characteristics** (vegetation- and soil-related properties) and processes that determine them, play a strong role in the predictability of near surface **atmospheric state variables**, NSASV, such as atmospheric temperature (T_{2m}) and relative humidity (RH_{2m}). Despite the implementation of a steady stream of improvements in the ECMWF Integrated Forecasting System (IFS) in recent years (Sandu et al., 2020; Boussetta et al., 2021) there are still considerable and persistent biases in LSSV and NSASV, especially for specific regions (e.g. the polar regions, including the third pole, and boreal regions), seasons (spring and autumn) and forecast ranges (beyond 2 weeks).

2. Motivation and broad scientific approach: (i) A significant **reduction of NSASV biases**, at a range of spatial and temporal scales (diurnal, monthly to seasonal, and climate scales), is an important motivation for the proposed research, for the benefit of Society through improved predictions and services. This requires maximum exploitation of the growing range of observations at ever-increasing spatio-temporal resolutions, including data obtained from sensors installed on airborne platforms (such as satellites) as well as in-situ data (from quality-controlled observatories). Continued and pro-active **Earth Observation-driven model system development** (Balsamo et al. 2018) of the coupled IFS Land Data Assimilation (DA) system (LDAS) and the ECLand model (Boussetta et al., 2021), will allow ECMWF and its collaborators to be prepared when new data from planned satellite missions become available. These kinds of developments form a central aspect of the ECMWF coupled land-atmosphere DA strategy (de Rosnay et al. 2022).

ii) A related motive is the notion that there is a strong **lack of realism** in the representation of **soil and vegetation** in Land Models (LMs), including in ECLand, which will impede the predictability of LSSV and NSASV. Vegetation parameters, such as leaf area index (LAI) and vegetation albedo (α_{veg}), are prescribed, and even if the CO_2 assimilation flux informs vegetation growth, the **interactions** between soil and vegetation are absent or poorly implemented. Moreover, important (interactions between) below-ground processes are missing (e.g. coupled heat- and water flow; Zeng et al., 2011). Additionally, soil parameters are **static**, defined once at model configuration stage (derived from globally distributed soil maps and empirical pedotransfer functions (PTFs), both of variable quality). PTFs link readily available soil properties (e.g. soil texture) to soil physical parameters (i.e. **hydraulic** parameters that determine *water* flow and storage) and **thermal** parameters (those that determine *heat* flow and storage), as well as to biogeochemical parameters for carbon/nutrient cycles (Van Looy et al., 2017). The choice and quality of the soil map and PTF leads to large uncertainties and affects model skill. Instead, here we propose the **timely** and **ambitious** implementation of a **fully dynamic soil-vegetation system in the IFS** (via novel developments of LDAS and ECLand), whereby the soil will be considered as a temporally variable medium that can be **'monitored' from above**. Within the IFS we will enable interaction between soil model parameters, their environmental and anthropogenic drivers, and soil and plant processes, by leveraging the new Multi-Parameter Regionalisation framework (MPR) available from the IFS cycle 49r1. Via a combination of EO and in-situ data, novel DA and mechanistic model equations we will obtain the **below-ground calibrated model properties**. We will largely focus on the dynamics of the **soil structure**, i.e., the arrangement of the solid soil particles and the pore space in between; here the pore-size *distribution* (PSD) is crucial. It is largely the PSD, and its spatio-temporal variability, that determines the *hydraulic* properties of soil. The mineral composition of soil particles (e.g. clay type) also has an effect, in particular in the context of the soil *thermal* properties (e.g. the thermal conductivity (λ) of quartz is 4 times larger than that of clay). Soil scientists and agronomists already increasingly view the **structure** of the soil system and its properties as temporally variable. Efforts are underway to reflect this notion in LMs (Fatichi et al., 2020), which we will build on via expert elicitation and literature surveys. Soil physical properties vary on sub-monthly to seasonal timescales due to land use and management activities, such as tillage and livestock trampling; freeze-thaw effects; vegetation growth (e.g. the beneficial effect of biopores on soil hydraulic properties, and hence on infiltration, percolation, and groundwater recharge) and fire: the charred soil (now hydrophobic) will have distinctly different hydro-dynamic and thermal properties. Finally, impacts of environmental change (e.g. melting of permafrost, and reduction of soil organic matter due to increased soil respiration) will also affect the soil hydro-thermal properties; these effects are important for climate-scale modelling (see Robinson et al., 2019). Key to these proposed developments (a dynamic soil-vegetation system with improved process descriptions, while making optimal use of EO and in-situ data via DA) is the need for a **paradigm shift** in the way we currently treat soil hydraulic and thermal theory in LMs. This requires a **unifying soil hydro-thermal theory**, whereby changes to soil structure (and related PSD) affect both hydraulic and thermal properties. This theory will have **soil matric potential** (Ψ_s) as the independent variable (Luo et al, 2022), rather than soil moisture content (θ), although they are linked via the water retention curve ($\theta(\Psi_s)$, a soil property that determines soil water storage and flow. The rationale is that mounting evidence suggests that accurate modelling (and monitoring) of the **soil-plant hydraulic continuum**, including Ψ_s-based plant water stress functions, will lead to improved prediction of LSSV and fluxes (Verhoef & Egea, 2014; Sabot et al., 2020; Wang et al., 2021). Also, reduction of soil hydraulic conductivity (K_s; it depends on Ψ_s, and directly determines water flow) is a primary driver of plant drought response (Carminati and Javaux, 2020): plants adapt root hydraulic conductivity (K_r) to match the soil conditions, to meet the atmospheric water demand (via changes to their leaf water potential, Ψ_l). Such regulations of root zone hydraulic properties, plant water status and transpiration can be reliably predicted by theory of the soil-plant hydraulic continuum, considering both above- and below-ground hydraulic traits as well as phenological and physiological parameters. **(iii)** In this context, we will explore the use of **'vegetation as a root-zone soil sensor' (VaaSS) for spatio-temporal derivation of subsurface properties**[1], using satellite observables that capture key aspects of the soil-plant continuum, and related water-energy and carbon exchanges, across near (NIR), thermal (TIR) and shortwave (SWIR) infrared, and microwave domains (MW) for determination of vegetation optical depth (VOD), vegetation water content (VWC), and Solar-Induced Fluorescence (SIF) (e.g. Konings et al., 2019, 2021). Fig. 1 provides a summary of the proposed approach. This diagram concerns the **STEMMUS-SCOPE** model, that combines a state-of-the-art soil physics model (STEMMUS: Zeng et al., 2013; Yu et al., 2018) with the SCOPE 'Soil Canopy Observation' model (Van der Tol et al., 2009). The ground-breaking approach described above will be implemented, tested and honed first with the STEMMUS-SCOPE model before we incorporate it into the ECLand-DA system, and subject it to detailed evaluation.

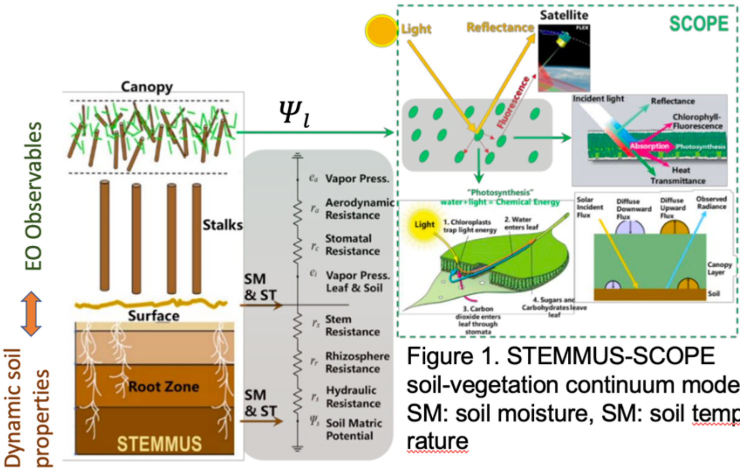

Figure 1. STEMMUS-SCOPE soil-vegetation continuum model. SM: soil moisture, SM: soil temperature

<u>3. Models, tools and datasets</u> These entail the STEMMUS-SCOPE model, the IFS system (LDAS and ECLand), in-situ data from PLUMBER2 & ESM-SnowMIP sites and observing networks (e.g. SYNOP, the International Soil Moisture Network) and aerial and satellite EO-data (VOD, VWC, and SIF). Focus will be on regions of interest (ROI) that will involve different types of L-A interactions and model bias types (e.g. the Tibetan plateau to study the effects of freeze-thaw and

[1] We will use a ML covariate-based GeoTransfer function (CoGTF) framework to translate near-surface soil properties into deep soil properties, accounting for key co-variates, such as vegetation type and climate.

soil-snow-atmosphere feedbacks on model biases (Krinner et al., 2018); agricultural bread-baskets for scrutiny of land use and management effects on model biases; and the Amazon for root-soil interactions and its implications for spatio-temporal variability of PSD, soil properties and related biases of LSSV and NSASV). ECLand will soon become open-source which will benefit the proposed project. Also, ECMWF is preparing an externalisation package that includes the Offline Surface Model and Analysis tool for easy compilation and running of the standalone ECLand versions for PLUMBER2 and ESM-SnowMIP sites. This package also includes 2D regional and global offline simulation configurations (as developed within the EartH2Observe project) for catchment evaluations at the spatial resolutions of 0.25° x 0.25°, 0.1° x 0.1° and 0.05° x 0.05°. We will also be employing the multiscale parameter regionalization system (MPR) which facilitates testing and calibration, and re-tuning/optimization of parameters.

4. Aims (A), Workpackages (WP) and Deliverables (D, also see Gantt chart in Section B):

A1: Implementation, testing and verification of the 'Vegetation as a root-zone Soil-Sensor' (VaaSS) approach using the STEMMUS-SCOPE model. **Team**: PDRA, Verhoef/Black (UoR), Zeng and University of Twente (UoT) colleagues, ECLand and LDAS ECMWF staff

WP1.1: Zeng's team is currently working on modularisation of STEMMUS-SCOPE as well as implementation of a DA system, which will facilitate coupling between STEMMUS-SCOPE and the Community Microwave Emission Modelling Platform (CMEM; de Rosnay et al., 2020) to develop a forward observation operator across the VIS-NIR-TIR-MW domains. To improve computing efficiency, a STEMMUS-SCOPE emulator has been developed based on physics-informed machine learning (PIML). A PIML-based CMEM emulator will be developed in our project, and seamlessly coupled with the STEMMUS-SCOPE emulator to construct an efficient forward observation operator for testing of the VaaSS approach. **WP1.2** Another key task of this WP concerns updating STEMMUS-SCOPE by implementing the **unifying soil hydro-thermal framework**. Central to this will be a set of hydraulic equations in which **PSD *directly* affects the hydraulic functions** (e.g. ploughing will destroy certain pore size classes and create new ones). Most LSMs use Van Genuchten hydraulic equations, which do not lend themselves to this approach, so an alternative approach is required (e.g., Kosugi's). The PSD, together with soil texture will also affect thermal conductivity (λ), which is currently modelled as an empirical equation that depends on θ, but increasingly theory is being developed with Ψ_{soil} as the independent variable (Luo et al, 2022), that we will build on. The new theory will be implemented in STEMMUS-SCOPE, and tested for a select set of locations, using combinations of EO observables (e.g. brightness temperature, backscatter coefficients, and higher level products like VWC, SIF and LST), and data from observatories detailed under Section 3. Next, this updated soil-plant model will be used to infer the dynamic state of vegetation (e.g. Ψ_l), and from that the state of the **below-ground part of** the **soil-plant hydraulic continuum** (largely derived from hydraulic connections between Ψ_s and Ψ_l , see Fig. 1. With that information the dynamic soil hydraulic and thermal functions can be constructed. The PDRA will use the STEMMUS-SCOPE DA system (see WP1.1) to tune model parameters using a perturbed physics approach. We will examine the introduction of new constraints, to avoid implausible future model states (so-called ecological dynamical constraints) and to explore model skill improvement via novel plant hydraulics **optimisation theory** (Sabot et al., 2020) within the DA framework. Note that the VaaSS approach will *also* yield continuous estimates of α_{veg}, LAI, and vegetation height (e.g. via VOD), from which aerodynamic roughness length can be derived, as well as SIF, see also WP2), so that phenology and the effect of land use changes will automatically be considered. **WP1.3** Using historical (remote and in-situ) observation and driving data, a step-wise testing of the effect of the updated theory and parameter estimates (compared to a Business as Usual (BAU) approach; i.e. standard soil maps, PTFs, standard hydro-thermal theory, and a fixed vegetation 'climatology') will be conducted for the ROIs, using suitable model-data comparison platforms and metrics (e.g. via GEWEX (Global Energy and Water Exchanges)' PLUMBER2 tool). We will also thoroughly explore **robustness of parameter sets and uncertainty**.

Deliverables D1.1 PIML-based CMEM-STEMMUS-SCOPE emulator. **D1.2.** Literature study on state-of-the-art soil hydro-thermal theory and VaaSS approaches. **D1.3** Expert elicitation workshops (e.g. via ISMC working groups, GEWEX GLASS SoilWat activity, and EO-DA forums). **D1.4** Updated CMEM-STEMMUS-SCOPE model/emulator (with theory derived from D1.2-1.3) and testing for ROIs **D1.5** Optimal set of model parameter distributions for the ROIs, and climatologies for vegetation parameters (e.g. LAI and α_{veg}), including associated uncertainties. **D1.6** Multi-year parameter sets for initial tests with IFS (see WP2), drawing from parameter distributions provided by D1.5. **D1.7** Papers to summarise the findings and recommendations from WP1.

A2: Extension of the ECMWF ECLand ensemble DA system to achieve novel dynamic soil-vegetation parameter estimation. **Team**: PDRA, ECLand coupled processes team led by Balsamo, ECMWF coupled DA team led by de Rosnay, Verhoef/Black (UoR), Zeng and UoT colleagues.

WP2.1: This task concerns replacing the soil-plant system theory in ECLand. Building on the outcomes of WP1 (including transferable soil/vegetation process modules), the **unifying soil hydro-thermal framework**, together with missing soil- and plant (hydraulic) processes, will be implemented in ECLand. As in WP1, step-wise offline testing, involving metrics assessing the **improvement of LSSV** prediction will take place, e.g. via the LANDVER validation package for land surface variables. We foresee that this will lead to improved thermal land-atmosphere coupling strength to address the current diurnal cycle biases and temporal shifts in T_{soil}, LST, and T_{2m}. **WP2.2**: Currently, ECMWF's DA system uses the latest observations of T_{2m} and R_{2m} (obtained from SYNOP), ASCAT and SMOS soil moisture products, SYNOP and national ground observations of snow depth, as well as NOAA/NESDIS snow cover information. Presently, ECMWF is updating ECLand by **refining and extending layer discretisation in the soil and snow profiles**, to improve predictions during drought, and of snow accumulation and melting. This has been shown to improve the link between the soil and the snow blanket (via base heat flow), and between the land surface and the atmosphere, resulting in an improved match of snow state variables and soil moisture content with satellite and ground observations[2]. Other developments are focused on **improving the model's vegetation phenology and its seasonality**, largely by implementing operational LAI, α_{veg} and VOD observation operators (see also WP1). This is important for the prediction of intra-annual (to improve predictability during spring and autumn) and inter-annual variability and representation of extremes (Boussetta et al. 2013). Whilst vegetation has a prominent role in warm seasons it can also interact with the snow duration therefore the evaluation will cover all seasons. In addition, a unified **offline** ensemble-based land DA system is being developed in ECLand. For this purpose, the development of ML-based observation operators that include soil hydro-thermal and vegetation parameters will be explored and coupled to the offline Ensemble ECLand DA system. We will **extend this system** to address **dynamic** soil-vegetation parameter estimation, allowing for assimilation of further EO observables across NIR-TIR-SWIR-MW domains (e.g. VOD, VWC, SIF, α_{veg}) using the **VaaSS** approach described under WP1. **WP2.3** Building on WP2.2, offline ensemble land DA experiments, analogous to the approach detailed in WP1, will be conducted using the MPR. As in WP1 we will be comparing to BAU, and intermediate scenarios (e.g. phenology derived from EO, but with the original non-dynamic soil system). In the final phase of this WP the enhanced coupled Ensemble LDAS suite will be used to derive regional to global time-series of soil and vegetation parameter fields. Land DA experiments and subsequent coupled land-atmosphere NWP experiments will be conducted to assess the impact of **improved dynamic soil-vegetation parameters** on NWP model skill, using standard ECMWF verification tools.

Deliverables D2.1 New version of ECLand with updated soil-plant hydraulic system theory and offline DA updates based on WP1. **D2.2** Production and evaluation of offline ECLand ensemble land DA experiments with updated soil & vegetation parameters. **D2.3** Production of medium-range coupled NWP experiments on selected case studies and assessment of updated parameters' impact. **D2.4** Papers presenting findings of D2.1-D2.3; communications to scientists and end-users. **D.2.5** Internal ECMWF report detailing the findings and recommendations.

5. Strategic importance This project firmly addresses the aims of AFESP's Theme 3 ("Data assimilation for the Earth System across a range of scales", specifically under Theme 3A: "Reduce observation gaps, particularly for ... the land- surface"; 3B: "Develop new and efficient observation operators that enable prediction systems to exploit the information content of observations that affect more than one ES component: ... land surface-hydrology"; 3C: "Maximising the Benefit of Observations... with a view to constrain (i) model state variables relevant to 1-month and/or kilometre-scale prediction, (ii) Model parameters (or processes) associated with model biases".

[2] Better VaaSS soil parameters found during snow-free periods will further improve year-round predictions.

It also contributes to Themes 1 & 2: the VaaSS approach offers a unique opportunity to radically improve our understanding (2A), process realism (2B) and simulation fidelity (2D) at the kilometre and monthly (1A) to sub-seasonal time scales (1C). The proposed ground-breaking (medium to high risk, high-reward) approach is compatible with AFESP's far-term timescales. While we are confident that this project will achieve most of its aims within its life span, more time is likely to be needed before a fully dynamic high-resolution Soil-Vegetation Model-Observing System will be operational at ECMWF. This project is an important investment towards this long-term goal. There are also large complementary benefits for existing services (e.g. CAMS (Copernicus Atmosphere Monitoring Service), C3S (Copernicus Climate Change Service) and projects such as Destination Earth (DestinE) and various Horizon Europe projects that ECMWF is currently coordinating. E.g., **CERISE** (2023-2026) aims to enhance the C3S portfolio quality, with the development of: (i) New and innovative coupled land-atmosphere data assimilation approaches, including **time varying vegetation** (ii) Land surface initialisation techniques. **CORSO** (2022-2025): CO2MVS (CO2 emissions Monitoring and Verification Support) Research on Supplementary Observations. The ECMWF Coupled DA Team will be exploiting satellite observations in coupled L-A assimilation to constrain vegetation water and carbon cycle variables. The **CoCO2** project (2021-2024) will also contribute to the development of the CO2MVS capacity, by delivering the prototype systems including Earth system models, DA techniques, and optimising the use of satellite and in-situ observations. A final anticipated strategic advantage derives from the fact that the proposed state-of-the-art LDAS system will make ECMWF and its partners extremely well placed to capitalise on the EO data from planned missions such as **CIMR** (2028-2033): All-weather high-resolution MW imagery of sea-ice and surface temperature and salinity covering polar regions; **CRISTAL** (2027-2034) to monitor sea-ice thickness and snow depth; **LSTM** (2029-2034: Land Surface Temperature Monitoring; **FLEX** (2025-2028; FLuorescence Explorer), a SIF mission.

References ***Balsamo***, G et al. (2018) Satellite and In Situ Observations for Advancing Global Earth Surface Modelling: A Review, doi: 10.3390/rs10122038; **Boussetta** S, *Balsamo* G... *de Rosnay*, P et al. (2021) ECLand: The ECMWF Land Surface Modelling System, doi: 10.3390/atmos12060723; **Boussetta**, S, *Balsamo*, G et al. (2013) Impact of a satellite-derived leaf area index monthly climatology in a global numerical weather prediction model, doi: 10.1080/01431161.2012.716543; **Carminati**, A & Javaux, M (2020) Soil Rather Than Xylem Vulnerability Controls Stomatal Response to Drought, doi:10.1016/j.tplants. 2020.04.003; ***de Rosnay***, P...*Balsamo*, G, et al. (2022) Coupled data assimilation at ECMWF: current status, challenges and future developments, doi: 10.1002/qj.4330; ***de Rosnay***, P et al. (2020) SMOS brightness temperature forward modelling and long term monitoring at ECMWF, doi: 10.1016/j.rse.2019. 111424; **Fatichi**, S et al. (2020) Soil structure is an important omission in Earth System Models, doi:10.1038/s41467-020-14411-z; **Hu**, Q, & Feng, S (2004) Why Has the Land Memory Changed?, doi:10.1175/1520 0442(2004)017<3236:WHTLMC>2.0.CO;2; **Konings**, A G et al. (2019). Macro to micro: microwave remote sensing of plant water content for physiology and ecology, doi:10.1111/nph.15808; **Konings**, A G et al., (2021) Detecting forest response to droughts with global observations of vegetation water content, doi:10.1111/GCB.15872; **Krinner**, G et al. (2018) ESM-SnowMIP: assessing snow models and quantifying snow-related climate feedbacks, doi:10.5194/gmd-11-5027-2018; **Luo S.** et al. (2022) Soil water potential: A historical perspective and recent breakthroughs, 10.1002/vzj2.20203; **Robinson**, D A et al. (2019) Global environmental changes impact soil hydraulic functions through biophysical feedbacks, doi:10.1111/GCB.14626; **Sabot**, M...*Verhoef*, A et al. (2020) Plant profit maximisation improves predictions of European forest responses to drought, doi: 10.1111/nph.16376; **Sandu** I, et al. (2020) Addressing near-surface forecast biases: outcomes of the ECMWF project 'Understanding uncertainties in surface atmosphere exchange', doi: 10.21957/wxjwsojvf; **Song**, Y et al. (2019). Soil moisture memory and its effect on the surface water and heat fluxes on seasonal and interannual time scales, doi: 10.1029/2019JD030893; **Van der Tol**, C...*Verhoef*, A, Su, Z (2009). An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance, doi: 10.5194/bg-6-3109-2009; **Van Looy** K.., *Verhoef* A et al. (2017) Pedotransfer functions in Earth system science: Challenges and perspectives, doi.org/10.1002/2017RG000581; ***Verhoef***, A & Egea, G (2014) Modelling plant transpiration under limited soil water: comparison of different plant and soil hydraulic parameterizations and preliminary implications for their use in land surface models, doi:10.1016/ j.agrformet.2014.02.009; **Wang**, Y, *Zeng*, Y, et al. (2021). Integrated modelling of canopy photosynthesis, fluorescence, and the transfer of energy, mass, and momentum in the soil–plant–atmosphere continuum (STEMMUS–SCOPEv1.0.0), doi:10.5194/gmd-14-1379-2021; **Yu**, L, *Zeng*, Y, et al. (2018) Liquid-Vapor-Air Flow in the Frozen Soil, doi: 10.1029/2018jd028502; ***Zeng***, Y et al. (2011) A simulation analysis of the advective effect on evaporation using a two-phase heat and mass flow model, doi:10.1029/2011WR010701; ***Zeng***, Y & Su, Z (2013) STEMMUS: Simultaneous Transfer of Energy, Mass and Momentum in Unsaturated Soil. (ITC-WRS Report). University of Twente, Faculty of Geo-Information and Earth Observation (ITC), Enschede, The Netherlands, pp. 6161–6164.

B. DESCRIPTION OF THE TEAM'S SKILLS AND CAPABILITY TO DELIVER

Team composition: The proposed programme of research will be tackled by a world-leading team of interdisciplinary experts on Land Modelling (LM), including soil physics and vegetation functioning, Earth Observing systems and data assimilation. They will make sure to recruit an exceptional and versatile researcher, with theoretical *and* applied scientific prowess. Principal Investigator (PI) **Prof. Anne Verhoef** (0.2 FTE) is Professor of Soil Physics and Micrometeorology at the University of Reading (UoR), UK, with extensive expertise in LM and field observation campaigns. She has a strong track record in the synergistic use of LM, field observations and remote sensing data. Over the years, she has built up a large portfolio (>£9M) of research grants broadly in the context of water-, food- and energy security, and related impacts of weather extremes (flooding, droughts, heatwaves), funded largely by UKRI. During her time at UoR (1999-) she has supervised 25 PhD students and mentored 15 post-doctoral researchers. She co-chairs the GEWEX Global Land Atmosphere System Study (GLASS) panel – since 2020, where she guides international developments relating to LM, and she is an executive board member of the International Soil Modelling Consortium – since 2017– where she co-leads (with Zeng) the soil thermal properties working group. Co-investigators (co-Is): **Prof. Emily Black** (0.1 FTE) is a senior scientist at the National Centre for Atmospheric Science (NCAS) and Professor of terrestrial processes and climate at the UoR. Her research focuses on land-atmosphere interactions, remote sensing, drought and climate services. She is an expert user of the JULES land-surface model and has been a Principal Investigator/scientific lead of several major projects (>£6M) including the recently completed NCAS Official Development Assistance (ODA) programme. **Dr. Patricia de Rosnay** (0.1 FTE) is Principal Scientist with the European Centre for Medium-Range Weather Forecasts (ECMWF) where she is leading the Coupled Assimilation Team in the Earth System Assimilation Section. She is an international leader in land data assimilation and coupled data assimilation. She is a member of the Scientific Steering Committee of the Center for Earth System Modeling and Prediction (CEMC) of CMA, Chair of the European Meteorological Society (EMS) awarding committee, and co-chair of the WMO Global Cryosphere Watch Snow Watch Expert Team. She is principal coordinator of the CERISE Horizon Europe project (2023-2026) as well as coordinator of the ECMWF contributions to the EUMETSAT H SAF and the ESA SMOS ESL projects. She has successfully mentored several PhD students, early career scientists and post-doctoral researchers and she has been a member of 18 PhD defense committees. She joined ECMWF in 2007 after previously working at CNRS at CESBIO in Toulouse from 2002 to 2007, and at IPSL in Paris from 1994 to 2002. **Prof. Gianpaolo Balsamo** (0.1 FTE) is an expert of Coupled Processes (CP) in Earth System Modelling at ECMWF. He has a PhD in Environmental Sciences and is an invited professor for Earth system science at Politecnico di Torino. He joined ECMWF in 2006 and he is now a principal scientist and team leader with his research focusing on Earth system coupled process modelling within the Integrated Forecasting System of ECMWF. He had previously worked for Environment & Climate Change Canada in Montréal, and for Météo-France in Toulouse, specialising on data assimilation of Earth Observations. **Dr. Yijian Zeng** (University of Twente, UoT, the Netherlands, 0.1 FTE) has extensive expertise in monitoring and predicting soil-plant-water-energy interactions, and mechanistic linking of above- and below-ground processes to satellite observables in the VIS-NIR-TIR-MW domains. He is the co-chair of the International Soil Modelling Consortium (ISMC) and leads the GEWEX GLASS SoilWat (Soil and Water) initiative, both aiming to integrate and advance soil system modelling, soil data gathering, and observational capabilities to improve the representation of soil and subsurface processes in Earth System Models. He is the PI of the EcoExtreML project (Accelerating Process Understanding for Ecosystem Functioning under Extreme Climates with Physics-Aware Machine Learning) funded by NLeScience Center, and is developing a Soil-Plant Digital Twin, consisting of the STEMMUS-SCOPE model, physics-aware machine learning algorithms, and a DA framework. Dr Zeng has supervised five PhD students and is currently overseeing 10 ongoing PhD projects. He has also mentored various PD researchers.

Project Management: The Project will be led by Verhoef in close collaboration with de Rosnay and Balsamo on the offline ECLand land data assimilation aspects, and with Zeng on the novel soil physical theory, and the development of the VaaSS approach via the Model Observing System based on STEMMUS-SCOPE. The Post-Doctoral Research Associate (PDRA) to be appointed will be based in the Department of Geography and Environmental Science (UoR), and advised by

Verhoef on boundary-pushing research on the soil-vegetation (hydraulic) continuum and terrestrial processes. He/she will have a hot desk in the Department of Meteorology to receive further advice on data-preparation, land-atmosphere interactions and their impact on weather and climate, and statistical climate metrics, from Black. To fulfil the tasks under WP1 the PDRA will spend 1-2 blocks of time at the UoT (the Netherlands) to receive training in the STEMMUS-SCOPE model and related DA systems. When working on the deliverables relating to WP2 the PDRA will spend time at ECMWF to work with the LDAS and CP modelling teams (led by de Rosnay and Balsamo, respectively, see Appendix 3 in part D). There will be quarterly project-team meetings, expert elicitation workshops (related to innovative soil physical theories and the VaaSS approach) and outreach events, and an annual meeting of the Project Advisory Board (PAB; to be established, featuring independent international experts in the land modelling, data assimilation and EO arena). A project kick-off workshop will take place in Month 1 for all researchers, Project Partners and PAB to agree the plan of work and DMP (Part C). The Gantt chart below details the approximate timing of the Deliverables outlined under Section 4 of part A (Description of the Proposed Research).

Deliverables	**Year and 6-month period** (1: January-June, 2: July-December)									
	2024		2025		2026		2027		2028	
	1	2	1	2	1	2	1	2	1	2
D1.1 CMEM-STEMMUS-SCOPE emulator	✓	✓								
D1.2. Literature study on state-of-the-art soil hydro-thermal theory and VaaSS approaches	✓	✓								
D1.3 Expert elicitation workshops		✓								
D1.4 Updated CMEM-STEMMUS-SCOPE model/emulator (with theory derived from D1.2-1.3) and testing for ROIs			✓	✓						
D1.5 Optimal set of model parameter distributions for the ROIs, and climatologies for vegetation parameters, including associated uncertainties					✓	✓				
D1.6 Multi-year parameter sets for initial tests with IFS; parameter distributions provided by D1.5						✓				
D1.7 Papers to summarise the findings and recommendations from WP1.			✓		✓					
D2.1 New version of ECLand with updated soil-plant hydraulic system theory and offline DA updates based on WP1.					✓	✓	✓			
D2.2 Production and evaluation of offline ECLand ensemble land DA experiments with updated soil & vegetation parameters							✓	✓		
D2.3 Production of medium-range coupled NWP experiments on selected case studies and assessment of updated parameters' impact									✓	
D2.4 Papers presenting findings of D2.1-D2.3; communications to scientists and end-users								✓	✓	✓
D.2.5 Internal ECMWF report detailing the findings and recommendations										✓

Risk management: The team has made sure that the outlined scientific approaches are feasible in principle, and that the high-risk aspects are managed carefully, as detailed below.
(1) ECLand is a modular portable code that provides a suitable and efficient environment to address implementation of new code, and novel soil-vegetation parameter estimation. The PDRA will be supported by both the Coupled Processes and the Coupled Assimilation Teams of ECMWF, led by experienced Team Leaders with extensive experience in project coordination and mentoring. (2) Developing new methodologies in NWP and Earth system models involves several risks related to the complexity of the system, difficulties to debug in a complex system, slow computing time, and related short testing periods which can hamper the evaluation process. The proposed approach to further develop **offline ECLand** will alleviate these difficulties and enable the PDRA to progress via a well-defined model environment that will allow them to fully focus on state-of-the-art soil-vegetation parameter estimation using EO and ground-based observables. The NWP impact will be assessed separately, following implementation in the coupled system, on ECMWF HPC platforms (3) The ambitious implementation of unified hydro-thermal theory in STEMMUS-SCOPE, and its further application in the OpenDA requires (i) a fully modularized model structure to enable efficient plug-in/out of certain process or parameterization modules; (ii) – excellent computing efficiency. We will manage the risks related to these requirements via the following on-going efforts. Firstly, STEMMUS-SCOPE is currently being re-developed to fit with good practice in Sustainable Software Development (in the EcoExtreML project, led by Y. Zeng). The model will be modularized further and equipped with Basic Model Interface (BMI) to facilitate its direct coupling with other models and DA frameworks (i.e., OpenDA). This effort will contribute directly to WP1 to facilitate the rapid development of a forward observation operator across the VIS-NIR-TIR-MW domain. Secondly, the UoT team has been applying physics-informed machine learning to develop a STEMMUS-SCOPE emulator, which will improve the computing efficiency tremendously, allowing for easy extension of the emulator as per WP1.1 under Section 4 of part A.

C. Technical Requirements and Data Management Plan

1. HPC requirements

The HPC requirements for this project are limited and will be provided in the form of in-kind support from SURF (see below) and ECMWF.

ECMWF: The PDRA will have an ECMWF user account that will give him/her access to the ECMWF HPC with standard RD user usage to submit and run experiments and use the ECFS and MARS (the ECMWF Meteorological Archive and Retrieval System) archives as well as the ECMWF tools. The use of the Offline LDAS and Offline Surface Model, as part of the ECMWF land surface platform, ECLand, will guarantee a good level of autonomy for the PDRA. The ECLand input will be provided by the ECMWF data centre archive accessible via MARS and via the Copernicus Climate Change Service (C3S) Climate Data Store (CDS). The data load is estimated to be 200 Mbyte for the in-situ simulations, that can run on a workstation in a matter of tens of minutes for the ESMsnowMIP sites and up to few hours for the PLUMBER2 sites. Gridded 2D simulation data load for the regions of interest will depend on the extent of the domain and the resolution considered and can reach up to several Gbyte. For 2D simulations the use of HPC facilities will be recommended so that the modelled and analysed output can be stored on MARS.

SURF: Extra HPC resources will be supported from SURF (National Computer Facilities of the Netherlands) at a yearly basis: 500,000 core-hour, 50 TB storage, 160 consultancy hours from Research Software Engineers. Supervision on the use of SURF will be provided by Dr. Zeng (who will be contributing 0.1 in-kind FTE to this project).

2. Additional data processing
Throughout the project, the ECMWF and SURF computational resources will be supplemented using the University of Reading cluster in order to facilitate collaboration with the UoR co-Is (Black and Verhoef). At this stage, we do not anticipate needing to use JASMIN because of the resource allocated from ECMWF and SURF. If JASMIN is required at a later stage of the project (for example to carry out supplementary model integrations), the resource can be provided via Co-I Black, who has access to the NCAS user space on JASMIN. The computational resource required for such integrations would be low and well within Co-I Black’s usage allowance.

3. Data archiving and management
The PDRA will undertake training to ensure that they are aware of data management protocols, including CF compliance and the need for sufficient meta-data. All data will be stored in netcdf format and will be CF compliant. During the project, all data will be archived at ECMWF on the MARS system. In accordance with journal policies, the data will be stored in such a way that it can be made available on request by the paper authors.

4. Code development and storage
All code developed during the project will be managed through a private project github repository. Code used in publications will be moved to a public repository, to comply with journal policies on reproducibility.

D. Appendix 3

Advancing the Frontiers of Earth System Prediction:
Summary of aligned Resource from ECMWF/MO/NCAS/other partners

Organisation	Named Staff	Assigned FTE	Technical contributions / Use of facilities	Other
ECMWF	**Patricia de Rosnay** Other LDAS members: Pinnington, Fairbairn, Garrigues, Weston, Herbert	0.1	Coordination of the land DA developments conducted in WP2. Coaching the recruited PDRA to work with the ECMWF ensemble land DA system.	The PDRA will take part in weekly Team meetings and have regular 1-2-1 meetings with the Team Leader. Organising scientific visits to work closely with the other LDAS team members (in-kind, no spare FTE).
ECMWF	**Gianpaolo Balsamo** Other CP members: Boussetta, Arduini, McNorton, Denissen, Choulga, Sutzl, Pedruzo, Weber	0.1	Coordinator of the surface coupled processes (CP) modelling work. Co-supervision of the recruited PDRA for the ECLand model parameters estimation/perturbat-ion.	Organising scientific visits to work closely with the CP team members. These are funded by external projects with already 1 FTE commitment, so their contributions will be in-kind
NCAS/UoR	**Emily Black**	0.1	Support for the analysis of climate data, including preparation of driving data files and selection of case studies	The team member is supported by the agreed 0.1 FTE in kind contribution from UoR
University of Twente	**Yijian Zeng**	0.1	Coordination of the development and extension of STEMMUS-SCOPE model. Co-supervision of the recruited post-doc on WP1 and WP2 Tasks.	UoT team members working with Zeng are funded by external projects with already 1 FTE commitment, and their contributions will be in-kind.

ECMWF | EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

14 July 2023 Tony.McNally@ecmwf.int

Professor Anne Verhoef
University of Reading
UK

To whom it may concern,

This letter is in support for the AFESP theme 3 proposal "Towards a high-fidelity Integrated Forecasting system via ground-breaking and ambitious data-assimilation of the dynamic soil-vegetation hydraulic continuum" led by Professor Anne Verhoef of University of Reading.

The proposal aims at improving the realism of soil and vegetation parameters, via calibration and enhancement of physical processes, and combining state-variables and parameter data assimilation. This scope fits well within ECMWF priorities and plans, it is supported by the existing ECLand modelling and data assimilation framework, and it will substantially expand its capabilities.

ECMWF commits 0.2 FTE equally partitioned between Dr. Patricia De Rosnay and Dr Gianpaolo Balsamo who will support the connectivity of the research within their respective teams.

In summary, we firmly believe that the above-mentioned proposal can have a strong impact on land and near surface weather monitoring, that it will be relevant for upcoming reanalyses and for environmental prediction.

Yours sincerely,

Tony McNally
Head of the Earth System Data Assimilation Section
Department of Research

UNIVERSITY OF TWENTE.

Anne Verhoef
Professor of Soil Physics and Micrometeorology
Department of Geography and Environmental Science
School of Archaeology, Geography and Environmental Sciences (SAGES)
The University of Reading
Whiteknights, PO Box 227, Reading, RG6 6AB, UK

FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

FROM:	DATE	PAGE
Dr. Yijian Zeng	15 July 2023	1/1
	OUR REFERENCE	
	WRS/YZ/2023-0715	

SUBJECT:
Statement of Collaboration and Support for the AFESP research project titled "Towards a High-Fidelity Integrated Forecasting System via Ground-breaking and Ambitious Data-Assimilation of the Dynamic Soil-Vegetation Hydraulic Continuum"

Dear Prof. Anne Verhoef

With this letter I pledge my strong hands-on support for the proposed AFESP project. I thoroughly enjoyed the process of helping you shape this proposal. If it gets funded it would be a unique opportunity to bring to fruition the ground-breaking ideas proposed in recent years during our GEWEX GLASS and International Soil Modelling Consortium (ISMC) meetings. I am hopeful that this funding call will allow us to build on our already extensive, albeit largely non-funded, collaboration on these important topics. With this we could make a real difference to the predictive skills of NWP and climate models.
The UoT (University of Twente) team will contribute to this project via a 0.1 FTE in-kind contribution of a senior scientist (mainly myself, and inputs from Prof. Bob Su), specifically with regards to:

- Work Package 1 (WP1) on the implementation, testing and verification of the "Vegetation-as-a-Soil-Sensor" approach using the STEMMUS-SCOPE model
- SURF (National Computer Facilities of the Netherlands) HPC resources at yearly basis: 500,000 core-hour, 50 TB storage, 160 consultancy hours from Research Software Engineers (in-kind)

I would be very happy to host and supervise the PDRA during key phases of WP1 to help him/her: 1. Understand the intricacies of the STEMMUS-SCOPE model; 2. Implement the novel state-of-the-art soil hydro-thermal theory and VaaSS approaches into STEMMUS-SCOPE; 3. to update the Physics-Informed Machine Learning based CMEM-STEMMUS-SCOPE emulator. I will also provide inputs into the (organisation of)the proposed expert elicitation workshops (I am co-chair of the ISMC and of the GEWEX GLASS SoilWat activity, and I am member of various EO-DA forums and advisor to the European Space Agency). I wish you the best of luck with your submission.

Best wishes,

UNIVERSITY OF TWENTE. FACULTY ITC

Dr. Yijian Zeng
Department of Water Resources
ITC Faculty of Geo-Information Science and Earth Observation, University Twente