**A. Project Information**

Project title: CHEESEHEAD19

Title of NSF award: Chequamegon Heterogeneous Ecosystem Energy-balance Study Enabled by a High-density Extensive Array of Detectors

NSF award number: 1822420

Project lead: Ankur Desai, University of Wisconsin – Madison, Department of Atmospheric and Oceanic Sciences

Submission date: March 23, 2020

**B. Overview of Project**

The Chequamegon Heterogeneous Ecosystem Energy-balance Study Enabled by a High-density Extensive Array of Detectors (CHEESEHEAD) is an intensive field-campaign designed to advance spatiotemporal scaling methods for heterogeneous land surface properties, fluxes and theories on the scales at which the lower atmosphere responds to surface heterogeneity. The overall goal is to investigate the atmospheric surface and boundary layer processes within the heterogeneous sub-grid of a single “grid” cell of a weather or climate model (i.e., 10x10 km) to improve representation of surface energy balance and atmospheric response to surface heterogeneity in these models.

The field campaign had three Intensive Observation Periods (IOPs) of one week each, spread across three months, to intensively sample the domain. We have a comprehensive dataset at hand that allows us to investigate the intensity of large-scale coherent structures and heterogeneity induced secondary circulations. However, due to the large range of scales of relevant atmospheric motions between 10-1 m and 104 m and three-dimensionality of the problem, the drivers and restrictions of the underlying processes can only be fully studied employing Large-Eddy Simulation (LES). This has mostly been done only for idealized conditions in the past (Inagaki et al. 2006; Huang et al. 2008). The temporally and spatially resolved data provided by LES furthermore allows to create space-time continuous land surface and lower atmospheric profile data that can be used to inform and fine tune machine learning approaches and subsequently model data comparisons. The observations conducted during CHEESEHEAD comprising of surface flux tower network, tall tower, ground-based remote sensing of the 3D-wind profiles and temperature and humidity distribution throughout the boundary layer, and a variety of ecological quantities, provide data at the lower and lateral boundaries of the simulation with sufficiently high spatial resolution and quality. Therefore, this data set is not only ideal for setting up an almost realistic LES but also for evaluating the LES output by comparing virtual measurements to tower and aircraft measurements.

**C. Science Objectives**

One proposed goal is to derive a **parametric heterogeneity correction** of dispersive fluxes missed by single-tower eddy-covariance measurementsfrom virtual towers within the LES. Therefore, we need realistic virtual tower measurements to investigate the energy balance non-closure problem by simulating the tower level turbulent characteristics as observed during the field campaign. The model data will be analyzed to come up with a parametric correction for heterogeneity effects. The advection and horizontal flux divergence can be quantified by a control volume analysis of the model output. Introducing virtual towers into the domain lets us compare model outputs with real observations for the same conditions and do interesting comparisons. For e.g. resolving the plant canopy lets us resolve the inertial sublayer impact on the data using the LES data. The parametric correction derived from virtual measurements will also be tested by applying it to CHEESEHEAD tower measurements.

We furthermore aim to **diagnose secondary circulations** and their relationship to surface heterogeneity scales and quantify the scale transport associated with them. Using 3-dimensional wind, temperature and water vapor LES data we plan to study the dynamical factors at play behind the physical processes by looking at vorticity budgets etc. Virtual flight will be setup during the run to mimic the actual IOP measurements, to sample the turbulent signals of wind, temperature and specific humidity. Flux footprint maps are being prepared from the actual flight observations; the LES run outputs will also be analyzed likewise and compared alongside these real-world measurements. The comparisons, along with spatio-temporally resolved LES outputs can help us investigate the role of landscape scale low frequency motions in surface-atmosphere exchange in line with our science objectives.

We also plan to setup virtual LiDAR measurements to replicate the actual LiDAR measurements during the measurement campaign which will help to interpret these observations in terms representativeness.

To **estimate the footprint of the virtual tower and flight measurements** within the simulated heterogeneous turbulent flow, we will use a Lagrangian approach where passive particles released at the surface are tracked during the simulation. This enables us to estimate time-dependent surface flux footprints (Steinfeld et al. 2008) at virtual sensors, which in turn are used to compare with projections for tower and aircraft data using the Environmental Response Functions machine learning approach (Metzger et al. 2013). Further, previous studies indicate that heterogeneous surface heating affect also the mixing processes at the boundary-layer top (Sühring et al. 2014), where warm and dry air is mixed from the above-lying free atmosphere into the boundary layer. This mixed air is transported downwards and can also affect near-surface observations (van de Boer et al. 2013). Also using Lagrangian particle tracking we will investigate of how much the virtual tower and flight measurements are affected by boundary-layer top mixing processes and how much uncertainty it introduces to the ERF-projected surface fluxes.

**D. Computational Experiments and Resource Requirements**

**D1. Numerical Approach**

The **Parallelized LES Model PALM** (Maronga et al. 2020, 2015, Raasch and Schröter 2001) will be used for numerical simulations. PALM solves the non-hydrostatic incompressible Boussinesq equations. To accurately simulate the physical processes as observed during the IOPs of the field experiment as realistic as possible, we will apply a **Land Surface Model (LSM)** with coupled soil, radiation and a **Plant Canopy Model (PCM)** with the topography for the domain. The use of the LSM and PCM runs instead of prescribed surface fluxes will enable the investigation of surface atmospheric feedbacks such as self-reinforcement of mesoscale circulations over the heterogeneous study domain (Wanner et al. 2019). This offers a detailed analysis of the physical processes involved and comparisons with observations such as computing radiation footprints for the surface energy balance, as well as, direct interaction to the synoptic and radiative forcing.  The inbuilt LSM implementation is based on the ECMWF-IFS land surface parametrization (H-TESSEL) and its adaptation in the DALES model (Heus et al. 2010). The canopy is modeled as a porous viscous medium that removes momentum from the flow (Shaw & Schumann, 1992; Watanabe, 2004), and acts as source/sink for heat, humidity, or passive scalar.

The LSM will be setup for each IOP test case, with land use classes and soil and vegetation data as measured/observed during the field experiment. Since the modeling domain is forested, the observed tree tops will be resolved using the inbuilPCM in PALM. The forest canopy will be initialized using observed parameters from canopy LiDAR measurements (Andresen et al. 2019) and field observations. In preparation for the realistic LES setup, our German CHEESEHEAD collaborators at the lab of Matthias Mauder performed idealized simulations using vertical grid nesting in combination with the inbuilt LSM and PCM (Wanner et al. 2019). For former projects, this group has successfully conducted high-resolution simulations of a large forested area (Kröniger et al. 2018) and contributed significantly to the development of an efficient two-way LES in LES nesting (Huq et al., 2019) in PALM.

**D2. Computational Experiments**

We propose to run LES case studies for two days per IOP using PALM with the models described in section D1. That way, we can cover changes in phenology between the three IOPs that are spread over summer and autumn, covering one week, respectively, as well as different atmospheric conditions. 8 ensemble runs will be performed for each studied day. For each IOP, a static driver, i.e. a netCDF file characterizing the soil, vegetation, plant canopy and domain topography (DEM), describing the lower boundary conditions will be set up.

To have realistic synoptic forcing data at the lateral boundaries of the model, PALM can be nested within a grid cell of WRF. In this case, the synoptic forcing data (temperature-, humidity and momentum advection, mean radiative forcing) will be derived from the WRF output. Since mesoscale models like WRF do not have resolved scale turbulent fluxes, we will need **adjustment zones** to allow for turbulence development at the inflow boundary so that we would have realistic turbulence within the domain of interest and the nesting approach does not affect the analysis anyway.  This approach works well with one-way nesting, where the child gets the boundary conditions from the parent with no feedback loop. A rough estimate of the required fetch length is about 3-5 times the CBL depth, which is of the order of 1 to 2 km.



Fig. 1: Schematic of the proposed nested grid setup

Fig. 1 shows a schematic presentation of the grid setup that we will use for the more realistic runs we want to conduct now. A rather coarse domain will cover a 36 by 36 km2 region to guarantee that turbulence is fully developed in the area of interest. A first nested domain with medium grid spacing will cover 27 by 27 km2 in which we will deploy virtual flight measurements similar to the airborne measurements conducted during the IOPs with flight legs measuring 25 km at 100 m and 400 m above ground. A second nested domain will measure 12 by 12 km2 to cover the CHEESEHEAD site where 20 eddy-covariance towers were deployed in the field. The resolution here has to be very high to enable comparisons between the eddy-covariance measurements, installed below 30 m above ground, and the virtual tower measurements in the LES. Therefore, we plan to include some additional sensitivity tests\* with different grid spacing.

**D3. Data output**

Virtual towers will be setup at the locations of towers on the CHEESEHEAD domain along with an additional tower grid (to investigate how well the actual tower network observes surface atmospheric exchanges). They will have multiple measurement heights at 20, 50,100 and 200 m to study how energy-balance closure and turbulent characteristics are resolved and to develop the parametric heterogeneity correction. Virtual flight measurements will be setup at 100 and 400 m above ground, similar to the field experiment, to sample turbulent measurements of wind, temperature and specific humidity.

We also plan to produce the surface and near surface XY cross section maps, to get space time continuous data (surface fluxes LST etc.) that can help inform sub-grid scaling algorithms. The model derived vertical profiles of humidity, temperature and flux divergences can inform additional questions about biomass heat storage, vertically dispersive fluxes etc. In addition to these, the fine resolution domain runs will be used to derive the high resolution XY maps of the CHEESEHEAD domain (10 x 10 km) at multiple heights, to calculate spatial fluxes etc.  The 3D volume data of winds, temperature and humidity will be written for the medium resolution domain and analyzed to look for secondary circulations as mentioned earlier.

**D4. Resource Requirements**

**Timing tests** were performed on Cheyenne for the model configurations described in section D2. The tests were done using an idealized Plant Canopy Model and Land Surface Model over a homogeneous domain with prescribed radiation (as described in Wanner et al. 2019) since the field campaign data is still being analyzed to come up with gridded spatial data for the domain.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Domain** | **lx \* ly \* lz (km)** | **dx \* dy \* dz (m)** | **nx \* ny \* nz** | **Number of PEs along** **X , Y** | **Total****PEs** | **Time step (s)** | **core hours for one hour of sim.** | **grid points total** |
| coarse | 36 \* 36 \* 6 | 100 \* 100 \* 60 | 360 \* 360 \* 100  | 10,18 | 180 | 20 | 3.35 | 12,960,000 |
| medium | 27\* 27 \* 1.8 | 30 \* 30 \* 20 | 900 \* 900 \* 90 | 15,12 | 180 | 6 | 63 | 72,900,000 |
| fine | 12 \* 12 \* 0.6 | 6 \* 6 \* 4 | 2400 \* 2400 \* 150 | 75,48 | 3600 | 0.65 | 7,020 | 864,000,000 |
| total |   |   |   |   |   |   | 7,086 | 949,860,000 |

**Overheads and scaling:** Since the domains will be nested, the parent domains will also run with the child domain time steps. We plan a final simulation time step of 0.4 seconds, giving us temporally high-resolution virtual measurements to compare with observed turbulence measurements. The coarse and medium run hours have to be up-scaled with the factor for change in time step. To optimize the computational demand, we will use a cubic grid spacing for the final simulations which will let us reduce the number of grid points in the vertical without losing too much of information. The fine resolution model will have a height of 300 m, that lets us resolve the surface layer turbulence and virtual tower turbulence measurements. The nesting is estimated to add 15% to the computational time (Hellsten et al., 2020, in preparation; Huq et al., 2019). The Lagrangian particle modeling won’t be turned on for the full duration and we estimate 20% of the total run time for it and data outputs would take 10%.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Domain** | **lx \* ly \* lz (km)** | **dx \* dy \* dz (m)** | **nx \* ny \* nz** | **Number of PEs along** **X , Y** | **Total****PEs** | **Time step (s)** | **core hours for one hour of sim.** | **Time step factor** |  **nz factor** | **Upscaled core hours per sim.** | **grid points total** |
| coarse | 36 \* 36 \* 6 | 100 \* 100 \* 100 | 360 \* 360 \* 60  | 10,18 | 180 | 0.4 | 3.35 | 50 | 0.60 | 100.5 | 7,776,000 |
| medium | 27\* 27 \* 1.8 | 30 \* 30 \* 30 | 900 \* 900 \* 60 | 15,12 | 180 | 0.4 | 63 | 15 | 0.67 | 630 | 48,600,000 |
| fine | 12 \* 12 \* 0.6 | 6 \* 6 \* 6 | 2400 \* 2400 \* 50 | 75,48 | 3600 | 0.4 | 7,020 | 1.63 | 0.33 | 3802.5 | 288,000,000 |
| total |   |   |   |   |   |   | 7,086 |  |  | 4,533 | 344,376,000 |

**Final Simulation estimates:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | Overhead Estimates |
| Model config. | IOPs (test cases) | Total Sim. hours | Sim. count | Upscaled Core hours per 1hr sim. | Total Core hours | Sensitivity\* (10%) | Nesting (15%) | Lagrangian particles (20%) | Data Output (10%) |
| coarse | 3 | 48 | 8 |  100.50  |  115,776 |   |  17,366.40 |   |   |
| medium | 3 | 48 | 8 |  630.00  |  725,760 |   |  108,864 |  145,152 |  72,576 |
| fine | 3 | 48 | 8 |  3802.50  |  4,380,480 |  438,048  |  657,072 |  876,096 |  438,048 |
| Total  |   |   |   |   |  5,222,016 |  438,048  |  783,302 |  1,021,248 |  510,624 |
| Total (rounded) |  |  |  |  |  7,975,300 |

Model output estimates are given below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model config. | Hourly Data Output (KB) | Hours of data written /day of sim. | No. of days | Archival Space (TB) |
| Fine |  18,004,383 |  18.00  |  48 | 15.56 |
| Medium |  12,177,872 |  18.00  |  48 | 10.52 |
| Coarse |  18,208 |  -  |  -  | 0 |
| Total (rounded) |   |   |   | 30 |

We estimate that we will need access to data for one of the IOP runs at a time, giving a project file space of 10 TB (1/3rd). Data analysis and visualization using Casper, 10,000 core hours.

Multi-Year Plan: We plan to complete the proposed simulations and data analysis by June 2021

**E. Data Management Plan**

Run outputs: The 3D volume data can become very storage intensive very quickly. We write this data only for the **medium** resolution domain. Ensemble averaging lets us filter out the random turbulence (primary circulations in the atmospheric boundary layer) and see coherent larger scale structures (secondary circulations) present over the larger scale domain (~ 30 kms). For this domain, we will also write the surface and first grid point XY cross section data to get spatio-temporal data for surface variables (fluxes, temperature, moisture etc.) and virtual measurements.

For the **fine** resolution domain, we will write the virtual tower data to compare with the actual tower network observations and XY cross-section data at multiple heights, needed for analysis such as spatial flux estimates (from covariances ).

The model output estimate table is based on these criteria. The domain averaged profile data, spectra etc. are not very heavy and for the ensemble runs are of the order of 1-2 GB.

The basic data management plan is as follows. Run outputs will be written to the scratch space initially. After each IOP run, the relevant 3d and cross-section data subset and the virtual measurement outputs for each IOP will be moved to HPSS archive for backup and intermediate storage.

Project files will be eventually copied to local disk space at UW- Madison for long term storage. All LES data will be made publicly available **within two months of production and analysis** from the NSF funded CyVerse Data Store (<https://www.cyverse.org/data-store>). Long-term archival data services will be provided by EOL data services. The CyVerse model outputs will also be linked to the NSF EOL CHEESEHEAD19 master data list and repository. Final, harmonized LES outputs and ERF scaled fluxes will be provided to EOL for the combined archive **by the CHEESEHEAD award end date (i.e., 31 July 2021).**

**F. Accomplishment Report**

The small-scale allocation associated with the same NSF award was used before the field experiment campaign to optimize measurement strategies. We conducted synthetic IOP runs forced by surface flux maps generated from ERF machine learning scaling algorithms. The runs were successfully performed for a 30 km x 30 km domain, with virtual towers and flight turbulent measurements by a graduate student at UW Madison, Sreenath Paleri, in collaboration with Dr. Matthias Sühring, a member of the PALM development team.

The model data was analyzed through the Environment Response Functions (machine learning based subgrid scaling approaches (Metzger et al. 2013)) to optimize the aircraft sampling patterns for the field experiment (addressing questions regarding flight orientation to mean wind, number of flight legs, start of a flight leg with respect to mean wind and domain etc.).

A manuscript detailing the numerical experiment design is under preparation:

S. Metzger, D. Durden, S. Paleri, M. Sühring, B. Butterworth, A. Desai, C. Florian, M. Mauder, L. Wanner, K. Xu: *Numerical experiment design doubles scientific return of surface-atmosphere synthesis*

**G. References**

Andresen, CG, May JA, Townsend PA, Desai AR (2019): Forest canopy structure characterization using high-density UAV LiDAR. Poster, AGU Fall Meeting 2019

Hellsten A, Ketelsen K, Raasch S, Maronga B, Sühring M, Knigge C, Barmpas F, Tsegas G, Auvinen M and Moussiopoulos N, Geosci. Model Dev., in preparation

Heus T, Van Heerwaarden CC, Jonker HJJ, Siebesma AP, Axelsen S, Dries K, Geoffroy O, Moene AF, Pino D, De Roode SR and Vilà-Guerau de Arellano J (2010): Formulation of the dutch atmospheric large-eddy simulation (dales) and overview of its applications. Geosci. Model Dev., 3: 415–444. <https://doi.org/10.5194/gmd-3-415-2010>

Huang J, Lee X and Patton EG (2008): A modelling study of flux imbalance and the influence of entrainment in the convective boundary layer. Boundary-Layer Meteorol., 127: 273–292. <https://doi.org/10.1007/s10546-007-9254-x>

Huq S, De Roo F, Raasch S and Mauder M (2019): Vertically nested LES for high-resolution simulation of the surface layer in PALM (version 5.0), Geosci. Model Dev., 12: 2523–2538. <https://doi.org/10.5194/gmd-12-2523-2019>

Inagaki A, Letzel MO, Raasch S and Kanda M (2006): Impact of surface heterogeneity on energy imbalance. J. Meteorol. Soc. Japan, 84: 187–198. <https://doi.org/10.2151/jmsj.84.187>

Kröniger K, De Roo F, Brugger P, Huq S, Banerjee T, Zinsser J, Rotenberg E, Yakir D, Rohatyn S and Mauder M (2018): Effect of Secondary Circulations on the Surface–Atmosphere Exchange of Energy at an Isolated Semi-arid Forest. Boundary-Layer Meteorol., 169: 209–232. <https://doi.org/10.1007/s10546-018-0370-6>

Maronga B, Gryschka M, Heinze R, Hoffmann F, Kanani-Sühring F, Keck M, Ketelsen K, Letzel MO, Sühring M and Raasch S (2015): The Parallelized Large-Eddy Simulation Model (PALM) version 4.0 for atmospheric and oceanic flows: Model formulation, recent developments, and future perspectives. Geosci. Model Dev., 8: 2515–2551. <https://doi.org/10.5194/gmd-8-2515-2015>

Maronga B, Banzhaf S, Burmeister C, Esch T, Forkel R, Fröhlich D, Fuka V, Gehrke KF, Geletič J, Giersch S, Gronemeier T, Groß G, Heldens W, Hellsten A, Hoffmann F, Inagaki A, Kadasch E, Kanani-Sühring F, Ketelsen K, Khan,BA, Knigge C, Knoop H, Krč P, Kurppa M, Maamari H. Matzarakis A, Mauder M, Pallasch M, Pavlik D, Pfafferott J, Resler J, Rissmann S, Russo E, Salim M, Schrempf M, Schwenkel J, Seckmeyer G, Schubert S, Sühring M, von Tils R, Vollmer L, Ward S, Witha B, Wurps H, Zeidler J and Raasch S (2020): Overview of the PALM model system 6.0. Geosci. Model Dev., 13, 1335–1372, <https://doi.org/10.5194/gmd-13-1335-2020>

Metzger S, Junkermann W, Mauder M, Butterbach-Bahl K, Trancón yWidemann B, Neidl F and Foken T (2013): Spatially explicit regionalization ofairborne flux measurements using environmental response functions. Biogeosciences, 10 (4): 2193–2217. <http://dx.doi.org/10.5194/bg-10-2193-2013>

Raasch S and Schröter M (2001): PALM - A large-eddy simulation model performing on massively parallel computers. Meteorol. Z., 10(5): 363–372, doi:[10.1127/0941-2948/2001/0010-0363](https://dx.doi.org/10.1127/0941-2948/2001/0010-0363)

Shaw RH and Schumann U (1992): Large-eddy simulation of turbulent flow above and within a forest, Boundary-Layer Meteorol., 61: 47-64. ​<https://doi.org/10.1007/BF02033994>.

Steinfeld G, Raasch S and Markkanen T (2008): Footprints in homogeneously and heterogeneously driven boundary layers derived from a lagrangian stochastic particle model embedded into large-eddy simulation. Boundary-Layer Meteorol., 129: 225–248. <https://doi.org/10.1007/s10546-008-9317-7>

Sühring M, Maronga B, Herbort F and Raasch S. (2014): On the Effect of Surface Heat-Flux Heterogeneities on the Mixed-Layer-Top Entrainment. Boundary-Layer Meteorol., 151: 531–556. <https://doi.org/10.1007/s10546-014-9913-7>

Van de Boer A, Moene AF, Schuttemeyer D and Graf A (2013): Sensitivity and uncertainty of analytical footprint models according to a combined natural tracer and ensemble approach. Agr. Forest Meteorol., 169: 1–11. <https://doi.org/10.1016/j.agrformet.2012.09.016>

Wanner L, De Roo F and Mauder M (2019): Optimizing Large-eddy Simulations for Investigating the Energy-balance Closure Problem at Typical Eddy-covariance Measurement Heights. Poster, AGU Fall Meeting 2019

Watanabe T (2004): Large-Eddy Simulation of Coherent Turbulence Structures Associated with Scalar Ramps Over Plant Canopies, Boundary-Layer Meteorol., 112: 307-341. ​[https://doi.org/10.1023/B:BOUN.0000027912.84492.54](https://doi.org/10.1023/B%3ABOUN.0000027912.84492.54).

**H. Figures and Captions**

Figure2: Maronga et al. 2015 (Figure 12)

Scalability of PALM 4.0 on the Cray XC40 supercomputer of HLRN. Simulations were performed with a computational grid of (a) 21603 and (b) 43203 grid points (Intel-Ivy Bridge CPUs). (a) shows data for up to 11 520 PEs with cache (red lines) and vector (blue lines) optimization and overlapping during the computation (FFT and tri-diagonal equation solver) enabled (dashed green lines). Measurement data are shown for the total CPU time (crosses), the prognostic equations (circles), and for the pressure solver (boxes). (b) shows data for up to 43 200 PEs and with both cache optimization and overlapping enabled. Measurement data are shown for the total CPU time (gray line), pressure solver (blue line), prognostic equations (red line), as well as the MPI calls MPI\_ALLTOALL (brown line) and MPI\_SENDRCV (purple line).