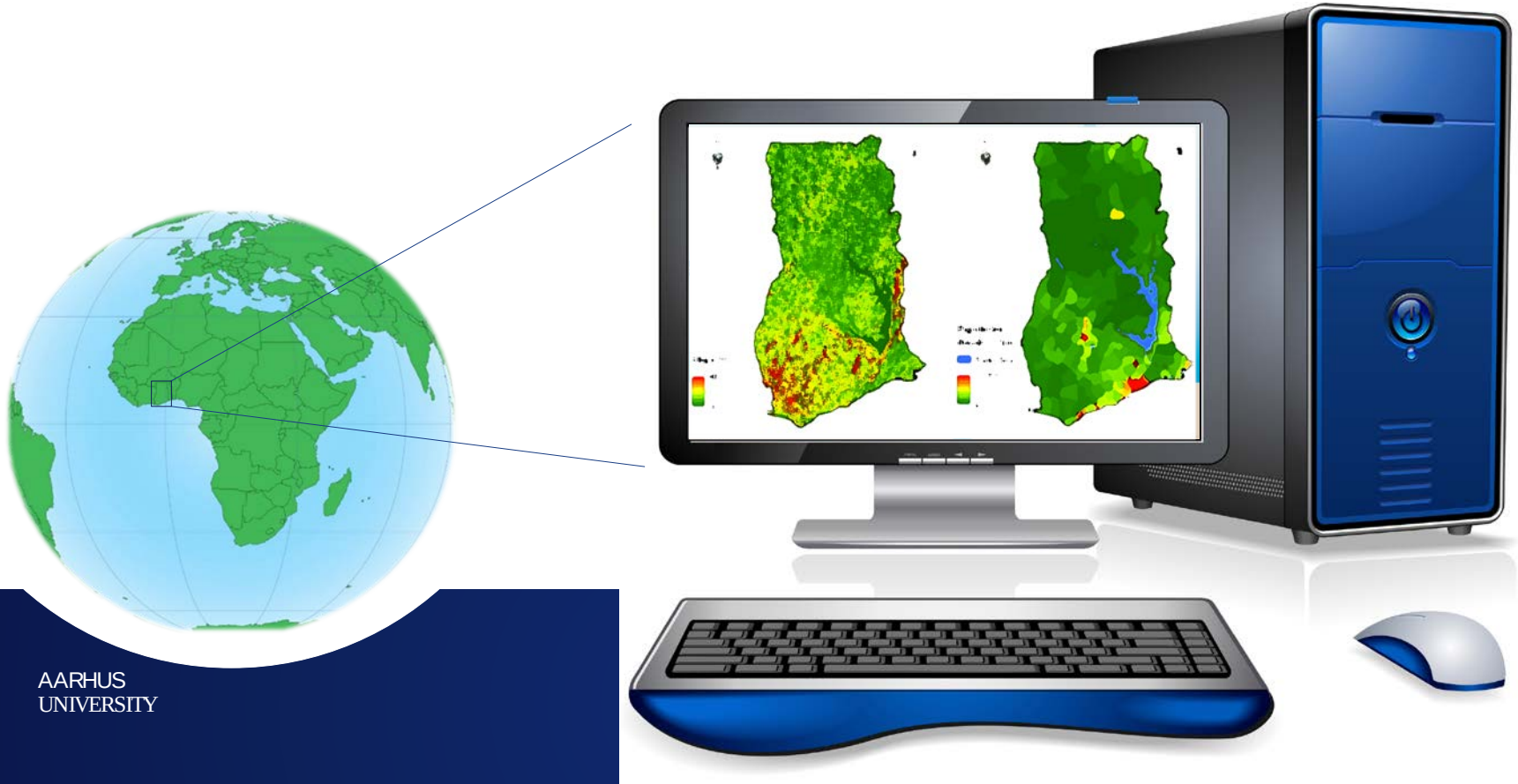
ANNUAL MEETING 2021 – STATUS FROM WP4: INTEGRATED ASSESSMENT AND SCENARIO BUILDING

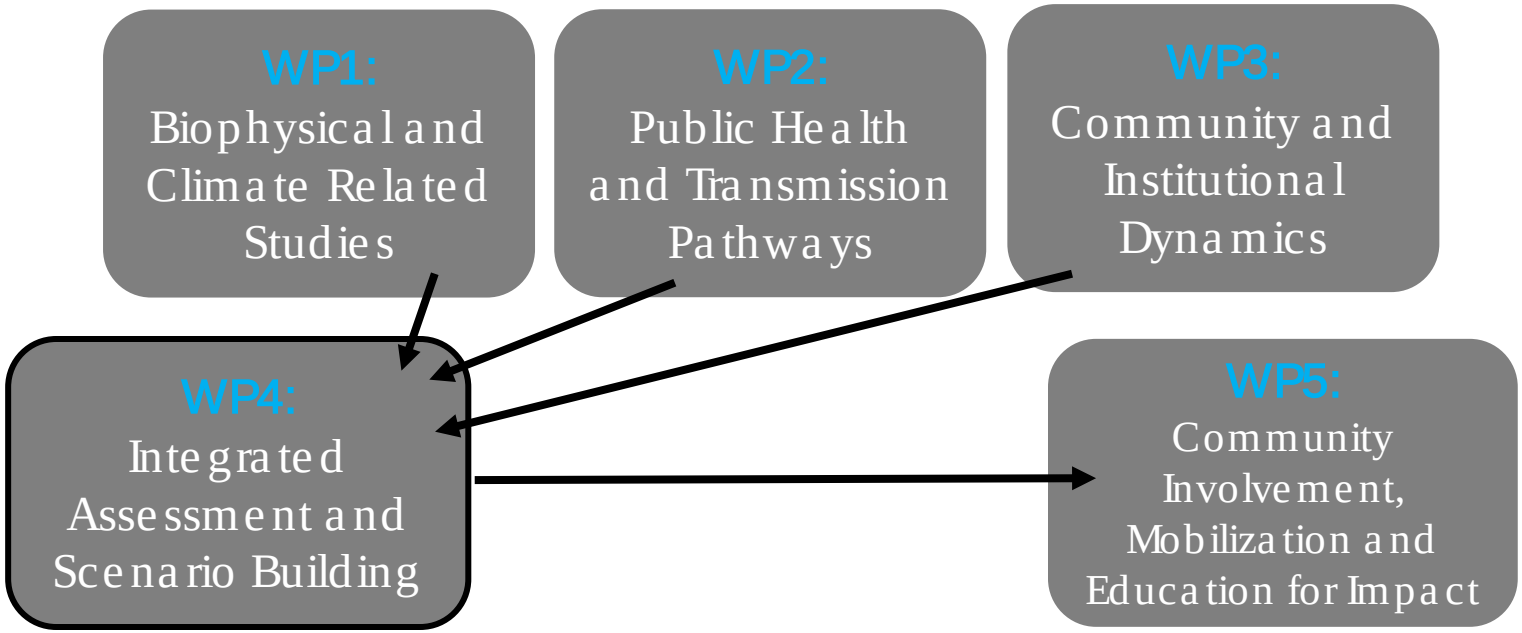
WP lead: Dennis Trolle (Prof., AU)
WP co-lead: Barnabas Amisigo (Research Scientist, UG)
Ph.D. student: Mr. Prosper Adiku
inputs from AU team and many other WP partners



WP4: Integrated Assessment and Scenario Building - overall goals

Overall goal of WP4 is to:

"generate statistical relations that estimate how key driving factors are influencing the occurrences of diarrhea, and further to build a software workflow around these relations that allow scenario simulations of diarrhea under future climate change and climate extremes."



SPECIFIC TASKS OF WP4

Task 4.1: Organize data from WPs 1, 2 and 3.
Task 4.2: Test an array of statistical approaches in terms of how well occurrence of diarrhea may be predicted from driving data provided by WPs 1, 2 and 3.
Task 4.3: Develop a Python-based **workflow**, which enable Monte Carlo simulations of scenarios through the most appropriate statistical model(s) from Task 4.2 (this will comprise the “probabilistic framework”).
Task 4.4: Develop Python-based **interactive visualizations** of scenarios (e.g. using Plotly), which can feed into dissemination activities in WP5.
Task 4.5: Deliver a training course on data analysis and applied statistics in Python.

TIMELINE OF WP4

WP	Tasks / activities	Partners	2020				2021				2022				2023				2024/25				
			3	6	9	12	3	6	9	12	3	6	9	12	3	6	9	12	3	6	9	12	3
WP4	1: Organize data from WPs 1, 2 and 3 into Python	1,2																					
	2: Test an array of statistical approaches for predicting diarrhoea	1,2																					
	3: Develop a Python-based scenario workflow	2																					
	4: Develop Python-based interactive visualizations	2																					
	5: Training course on data analysis and applied statistics in Python	2																					

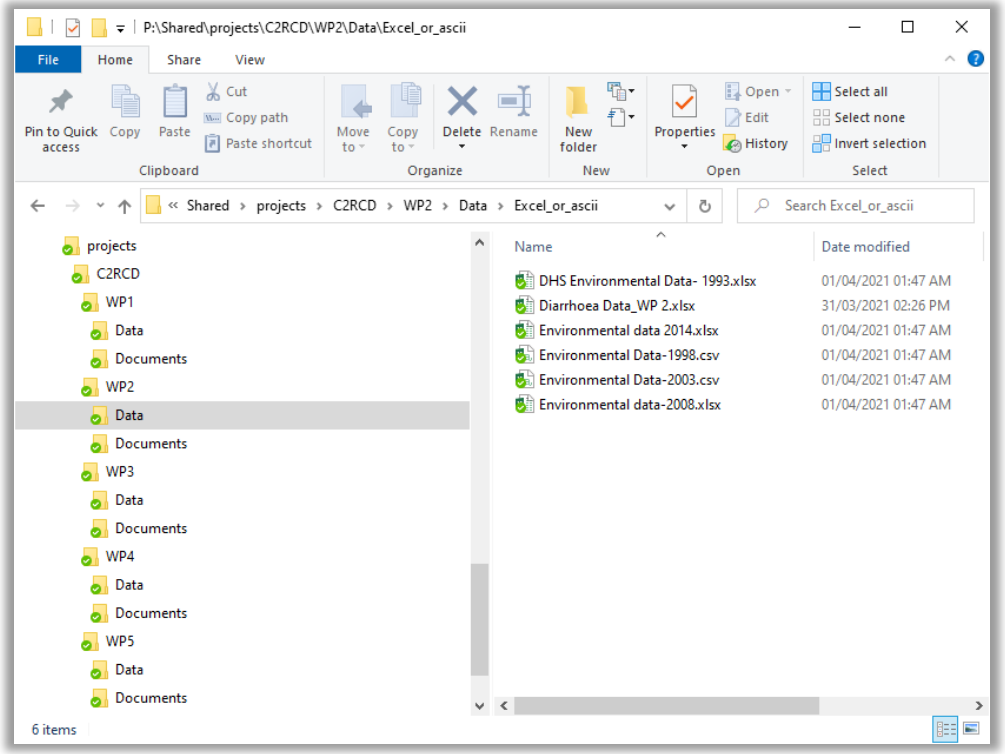
Partners		Colour code	
1	University of Ghana (Ghana)		Upstart
2	Aarhus University (Denmark)		Ongoing activity/input to other WP's
3	Peoples Dialogue (Ghana)		Scientific papers and dissemination
4	EPA (Ghana)		Workshops/trainings

Task 4.1: Organize data from WPs 1, 2 and 3 into Python compliant data structures.

- **Output:** Project's data storage (data from WPs 1, 2 and 3 is collected and organized electronically)
- Enable data documentation, sharing and backup
- Example of expected data formats:
  - GIS (vector, raster, dbf)
  - csv, excel

Task 4.1: Organize data from WPs 1, 2 and 3 into Python compliant data structures.

- **Status after year 1:**
- Different ways of organizing data have been assessed
- **pcloud** chosen as most suitable solution to share and backup project data (up to 10 GB free data)
- Folder structure and initial data already in place

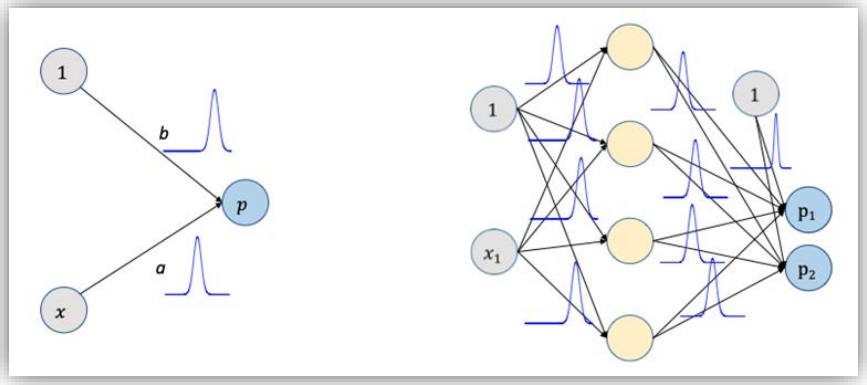


- **Plan for year 2:**
- Circulate quick-guide on how to get started with pcloud
- Send invite to projects folder structure
- Start working with data

WP	Tasks / activities	Partners	2020				2021				2022				2023				2024/25				
			3	6	9	12	3	6	9	12	3	6	9	12	3	6	9	12	3	6	9	12	3
WP4	1: Organize data from WPs 1, 2 and 3 into Python	1,2																					

Task 4.2: Test an array of statistical approaches in terms of how well occurrence of diarrhea may be predicted from driving data provided by WPs 1, 2 and 3.

- **Output:** Evaluation of an array of data analysis and statistical tests for predicting incidence of diarrhoea
- Learn from past examples
- Once data from WPs 1, 2 and 3 comes in start various tests and analysis
- Examples of approaches to be tested:
  - linear regressions, additive models
  - decision trees
  - Bayesian and neural networks
  - structural equation modelling, geostatistical approaches
- **Status after year 1:** state-of-the-art has been reviewed
- **Plan for year 2:** conduct test runs of statistical tests based on incoming data (and use collected data and tests in course preparations in Task 4.5 where possible)



WP	Tasks / activities	Partners	2020				2021				2022				2023				2024/25				
			3	6	9	12	3	6	9	12	3	6	9	12	3	6	9	12	3	6	9	12	3
	2: Test an array of statistical approaches for predicting diarrhoea	1,2																					

THE POTENTIAL OUTCOME…

- Geostatistical model to predict diarrhea prevalence in selected areas in Ghana, using the stochastic partial differential equation (SPDE) approach and the integrated nested Laplace approximation (INLA).

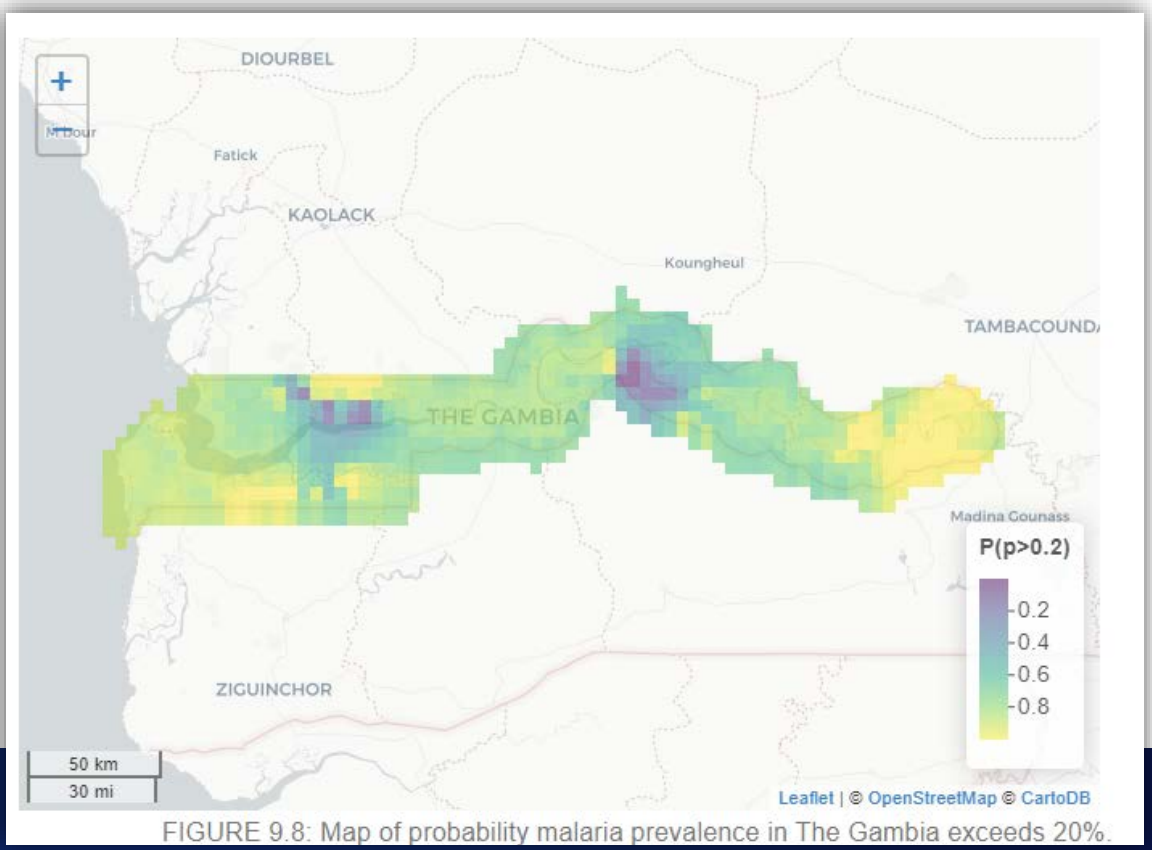


FIGURE 9.8: Map of probability malaria prevalence in The Gambia exceeds 20%.

Moraga, Paula. (2019). Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny. Chapman & Hall/CRC Biostatistics Series. https://www.paulamoraga.com/book-geospatial/

Task 4.3: Develop a Python-based workflow, which enable Monte Carlo simulations of scenarios through the most appropriate statistical model(s) from Task 4.2 (this will comprise the “probabilistic framework”).

- **Output:** Development of a workflow, comprising a series of Python (or R) script(s), which can effectively comprise and execute the workflow for predicting diarrhea prevalence
- Code/scripts will be shared via pcloud along with the data
- **Status:** planned to begin by end of year 3

WP	Tasks / activities	Partners	2020				2021				2022				2023				2024/25				
			3	6	9	12	3	6	9	12	3	6	9	12	3	6	9	12	3	6	9	12	3
	3: Develop a Python-based scenario workflow	2																					

Task 4.4: Develop Python-based interactive visualizations of scenarios (e.g. using Plotly), which can feed into dissemination activities in WP5.

- **Output:** Development of visualization routines in Python (or R) for dissemination of scenario results
- Should be presented on maps (potentially on project's website), and allow some sort of interaction (e.g. panning map, zooming, selection between different scenario results)
- Code/scripts will be shared via pcloud along with scenario results
- **Status:** planned to begin in year 4

WP	Tasks / activities	Partners	2020				2021				2022				2023				2024/25				
			3	6	9	12	3	6	9	12	3	6	9	12	3	6	9	12	3	6	9	12	3
	4: Develop Python-based interactive visualizations	2																					

Task 4.5: Deliver a training course on data analysis and applied statistics in Python.

- **Output:**
- Original plan:
  - Organize and host a training workshop at UG (in 2022) in **year 3**
- Updated plan:
  - Develop a completely online introduction course, e.g, "intro to Python and applied data analysis and statistics" (allowing people to work with this in their own time – but also schedule a live follow-up event) – this course will serve as an introduction to a more advanced course, which is planned to take place at UG in 2022 (online course will be developed in **year 2**, and will be available for all throughout the projects lifetime)
  - Organize and host a training workshop at UG (in 2022) in **year 3**

WP	Tasks / activities	Partners	2020				2021				2022				2023				2024/25				
			3	6	9	12	3	6	9	12	3	6	9	12	3	6	9	12	3	6	9	12	3
	5: Training course on data analysis and applied statistics in Python	2																					

SUMMARY OF WP4

WP	Tasks / activities	Partners	2020				2021				2022				2023				2024/25				
			3	6	9	12	3	6	9	12	3	6	9	12	3	6	9	12	3	6	9	12	3
WP4	1: Organize data from WPs 1, 2 and 3 into Python	1,2																					
	2: Test an array of statistical approaches for predicting diarrhoea	1,2																					
	3: Develop a Python-based scenario workflow	2																					
	4: Develop Python-based interactive visualizations	2																					
	5: Training course on data analysis and applied statistics in Python	2																					

Overall, progress follows the planned timeline

Activity level in WP4 generally increases in year 2 (as Task 4.2 kicks off)

Specific action items for 2021:

- Continuously populate the projects data repository (Task 4.1)
- Initiate statistical test (Task 4.2)
- Develop course(s) based on a combination of online training and the physical training (physical training planned for 2022) (Task 4.5)
- At next annual meeting we hope to be able to show first statistical results based on data collected in project

THANK YOU.