UWC West

Technical Annex, a feasibility study

This document is part of the Memorandum of Understanding
between the Danish Meteorological Institute, Icelandic Meteorological Office, Met Éireann and Royal Netherlands Meteorological Institute
concerning the preparation towards a joint operational exploitation of science-based short-term Numerical Weather Prediction system including ensemble prediction from 1st of January 2023 and the availability of a suitable HPC infrastructure from the end of the second quarter of 2022

<empty page>

Table of Contents

<empty page>

Introduction

As stated in the accompanying MOU, the main purpose of this collaboration is to enable the Parties to accomplish joint operational exploitation of science-based short-term Numerical Weather Prediction system including ensemble prediction from 1st of January 2023 and the availability of a suitable HPC infrastructure from the end of the second quarter of 2022.

This document details the outcome of a technical feasibility study assessing the NWP model configuration and domain sizing that would be achievable given a planned IT infrastructure and an expected budgetary envelope.

Executive Summary

It has been established that for an expected budgetary envelop of €2.8m[1] per annum an HPC facility consisting of two HPC systems with a shared storage subsystem is most feasible.

The location of the HPC facility should be determined during the procurement process on the basis of a feasibility study with regard to minimising operational costs associated with energy and as well as costs borne by each of the Parties related to data communications. This location should be based in one of the countries of the Parties.

It is planned that this HPC facility will support an operational common NWP production system by 1 Jan 2023. The Harmonie-Arome canonical configuration of the shared ALADIN-HIRLAM NWP System will be used. Two primary domains will be used to produce NWP forecasts for UWC-West; a Western European domain and a Greenland domain. Furthermore, a number of additional domains and downstream operational models will be exploited by the Parties individually.

There is uncertainty in estimating the HPC resources that can be achieved in 2022 with the expected budgetary envelope. Possible model and domain configurations have been explored for a range of HPC resources achievable in pessimistic, realistic and optimistic scenarios. At a minimum under the pessimistic scenario, it is expected that over the Western European domain an ensemble prediction system at 2.5km horizontal resolution with 18 ensemble members will be achieved.

It is planned to implement NWP configurations which maximize the use of the available HPC capacity and optimise the skill of the forecast data produced. The input of the pooled resource of UWC scientists and technicians as well as scientific developments in the wider ALADIN-HIRLAM community will inform the decisions of optimum model configuration.

[1] Based on the current (2017) spending of KNMI and DMI and the expected available budget for Met Éireann and IMO in 2022

HPC Facility Feasibility Study

The following four alternative technical solutions for a HPC facility were considered:

1. A single HPC and storage system at one location. This solution provides most flexibility but offers no resilience in addition to that of the HPC and storage system.

2. Two HPC systems with a shared storage system on one location. The HPC systems are located in different machine halls and have separate power supplies. This adds an extra level of resilience since, if one HPC system fails, production can be moved to the other system.

3. Two HPC systems with each a dedicated storage system on a single site. Both systems are installed in different machine halls with a separate power supply. In this case production can also continue when one of the storage systems fails.

4. Two HPC and storage systems on two geographically distant locations. This solution requires that the infrastructure at two sites is prepared for the installation of the HPC facilities. This solution offers also geo-redundancy.

The relative estimated costs of the above solutions is given in the table below, where X is the cost of solution option 1.

Technical solution	Relative costings
1	X
2	1.2X
3	1.3X
4	1.6X

From a costing analysis from 2017 it is estimated that technical option 1, delivering a capacity of 7 M-units[2] would have a total cost of ownership[3] of €5.81m per annum, assuming a lifetime of 5 years.

It is envisaged that the purchasing power will increase by 2022 and for the same investment in euro the affordable capacity will increase. Growth factors[4] of 2 (pessimistic), 2.5 (realistic) and 3 (optimistic) were considered.

[2] M-units are a measure of HPC capacity based on the UWC East (April 2017) configuration. See section Definitions for more details.

[3] TCO (total cost of ownership). This includes data centre, hardware, maintenance, running costs related to electricity and cooling, and exludes IT operations and data communication costs. Per annum costs assume a 5 year lifetime.

The estimated annual costs of a 7 M-unit HPC facility for each of the 4 technical solutions is detailed in the table below, for the pessimistic, realistic and optimistic scenarios.

Technical solution	Annual costs (€m/year)		
	Optimistic	Realistic	Pessimistic
1	1.94	2.32	2.91
2	2.32	2.79	3.48
3	2.52	3.02	3.78
4	3.10	3.72	4.65

The first of the four technical solutions is not recommended since it does not provide a sufficient operational resilience. Technical solution #2 gives an acceptable level of operational resilience and should for that reason be preferred over solution #1. The expected budgetary envelope in a pessimistic scenario will beinsufficient to purchase a 7M unit HPC system. However, in that case reduction of the operational capacity is recommended over compromising on operational resilience.

In a realistic scenario, a 7 M-unit system is feasible for technical solution #2. It is also assessed that this technical solution offers sufficient operational resilience. DMI/IMO are running this configuration already without any difficulties.

In the optimistic scenario, the third technical solution, i.e. a HPC facility with two separate storage subsystems, becomes financially feasible. The fourth, geo-redundant solution is not financially feasible under any of the scenarios identified.

Given the results of the feasibility study summarised above it is planned to acquire a HPC Facility consisting of two HPC systems with a shared storage subsystem, to maximize value for money. During the procurement phase of the project, the existing in-house expertise of the Parties related to acquiring, integrating and operating HPC facility will be used to maximum benefit.

The location of the HPC facility will be determined during the procurement process on the basis of a feasibility study with regard to minimising operational costs associated with energy and costs borne by each of the Parties related to data communications. The HPC facility should be located in one of the countries of the Parties.

Model and Domain Configuration Feasibility Study

As a principle each of the Parties will receive an equitable return on investment from the HPC system.
The current model domains (May 2018) of the Parties are shown on a map in figure 1 (left). Two possible common domains which have been considered in the feasibility study are shown overlying the current domains in figure 1 (right).

[4] See section Definitions for more details.

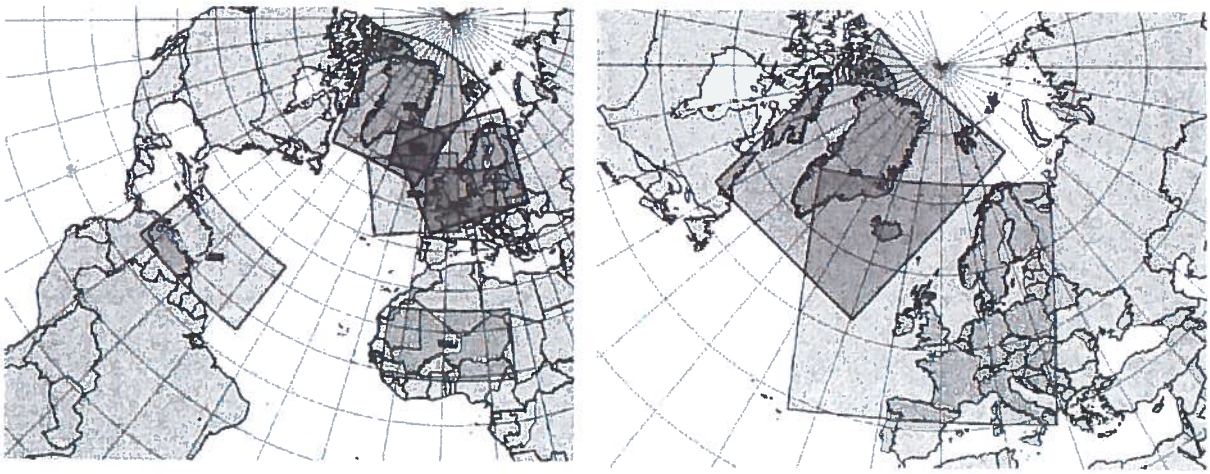

Figure. 1: Current operational domains of UWC West (left, May 2018) with Danish domains in red, Icelandic domain in blue, Netherland domains in orange and Irish domain in green. A possible two-domain setup is shown in the right panel.

A feasible set-up consistent with a 7 M-unit system in a realistic scenario is shown below. The Harmonie-Arome canonical configuration. The Harmonie-Arome canonical configuration of the shared ALADIN-HIRLAM NWP System will be used. Two primary domains will be used to produce NWP forecasts for UWC-West; a Western European domain and a Greenland domain. Other domains included are a Caribean operational domain operated by KNMI, a selection of high-resolution nested model configurations, and a number of downstream applications such as wave/ocean and air quality models. Further details are listed in the definition section. This setup is indicative of a feasible operational implementation in a realistic scenario, but can be adapted to future requirements, by adjusting ensemble size, domain size, resolution or forecast length.

DOMAIN NAME	Cost [M]
Main European domain (2.5 km, 6x3 members)	2.29
Greenland domain (2 km, 2 members)	1.24
Dutch Caribean (2.5 km, 1 member)	0.16
High resolution nested models (x3)	1.23
High resolution nested model, Dublin implementation	1.00
Downstream operational applications	0.70
Non allocated	0.38
Total	7.00

Definitions

<M-unit>

Applied reference model for computing power (UWC East M-units).
The following formula is used to determine the number of operations needed to complete an NWP ensemble run

#operations = (NLON x NLAT)/Gsize x VLEV x FL/TS x ENS

Where NLON and NLAT respectively represent the size of the model domain in the longitude and latitude, Gsize is the grid size, VLEV is the number of model levels, FL is the length of the forecast, TS (time step) is the time interval between two consecutive model iterations and ENS is the number of members in the ensemble. The UWC East (April 2017) configuration is used as normalized reference value (1 M-unit) for computing power.

	NLON (km)	NLAT (km)	VLEV	Gsize (km)	FL (hr)	TS (sec)	ENS
UWC East	750	960	65	2,5	56,4	75	10

<model characteristics>

DOMAIN NAME	NLON	NLAT	NLEV	FL	TS	ENS
Main European domain	1440	1440	90	54	75	18
Greenland domain	1500	1800	90	54	60	2
Dutch Caribean	1200	800	90	24	30	1
Mali	1200	800	90	24	30	1
High resolution nested models	1000	1000	90	24	30	1
High resolution nested Model, Dublin implimentation	1500	1500	90	24	30	1

<HPC performance per €>

Based on the most recent KNMI HPC experiences, the price-performance ratio hpc improved by a factor of 4 in 6 years (2010-2017). We also see a similar trend in the literature when analyzing the price-performance ratio of processors over a longer period. If we use these figures as a starting point, a growth factor in performance between 2 and 3 with a constant budget is a realistic estimate.