Coupling at scale: Experiences from the INCITE project

Kristian Mogensen, Inna Polichtchouk, Ioan Hadade, Sam Hatfield,
Nils Wedi (all ECMWF) and Valentine Anantharaj (ORNL)

ECMWF

Overview of talk:

- Overview of the INCITE project
- The ECMWF coupled modelling system
- The coupled model setup use for our INCITE
  - ECMWF ATOS
  - DOE Summit
- Scientific and technical challenges
- Some preliminary scientific results
- Summary and outlook

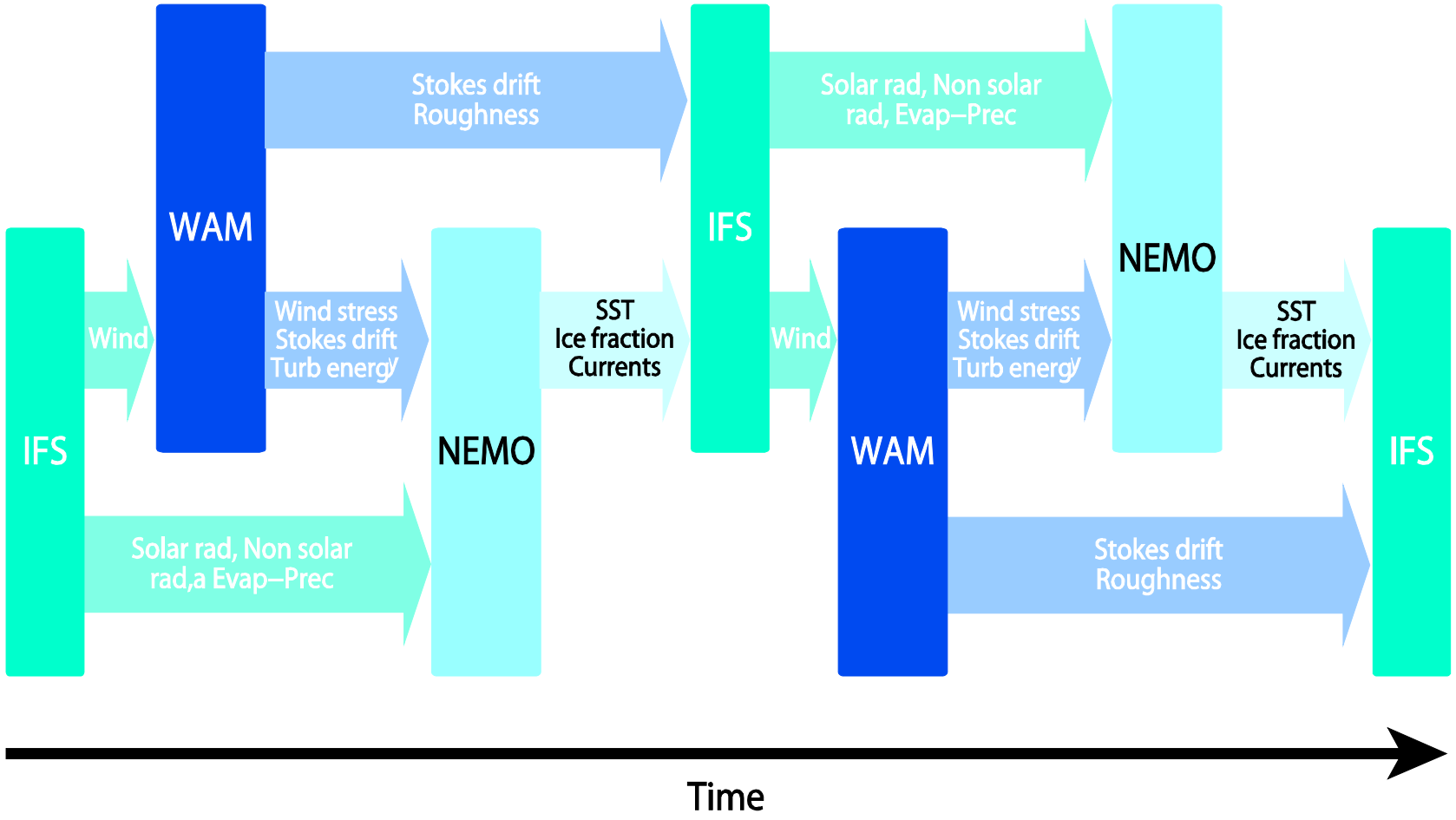
Overview of the INCITE project

- The Innovative and Novel Computational Impact on Theory and Experiment (INCITE) is a US DOE program to allocate resources on DOE HPC systems
- Nils Wedi *et al* has previously used Summit to perform 1.4 km uncoupled atmospheric only simulations under the INCITE 2020 allocation
  - https://doi.org/10.1029/2020MS002192
- In 2022 we were allocated resources on Summit to perform coupled simulations with high resolutions for both atmosphere and ocean
- Scientific focus was on the atmosphere-ocean interaction tropical cyclone conditions during 4 hurricanes: Irma (2017), Florence (2018), Teddy (2020) and Ida (2021)
  - Cases selected based on availability of ALAMO floats measuring ocean temperature and salinity with high temporal frequency
- At the same time, we were installing 4 new ATOS clusters in our new ECMWF data center in Bologna Italy
  - Access to a single cluster during the installation phase for about a week

The ECMWF coupled model

- Coupled modelling system consisting of:
  - IFS: Atmospheric model developed jointly with MeteoFrance (who calls it ARPEGE)
  - ECWAM: Wave model based on the original WAM model
  - NEMO: External ocean model
    - 3.4.1 used in our operational setup with ORCA025
    - 4.0 is being introduced and was used in this study with eORCA025 and eORCA12
- Operational models
  - HRES CY47R3 9 km deterministic 10 days twice daily
  - ENS CY47R3 18 km 51 members 15 days twice daily
    - Resolution upgrade to 9 km expected next year
  - Monthly CY47R3 36 km 51 member 46 days Monday and Thursday 0z
    - Starts from ENS
    - New configuration is under consideration
  - SEAS5 CY43R1 (November 2017) 36 km 7 months every month, 13 months every 3 months

Schematic overview of coupled system:



The LIM2/SI3 models needs additional fields from the IFS/WAM

Current ECMWF coupled system:

- Design principle: the atmosphere and wave models don't know anything about the inner workings of NEMO
  - There is a coupling layer, which have access to all NEMO F90 modules *(e.g.* data), but only accepts data from IFS/WAM as arguments
    - Coupling fields to/from NEMO is passed as subroutine arguments to this layer in IFS/WAM
    - Grid information from IFS/WAM needs to passed as arguments as well
- All regridding is done within the interface layer:
  - Interpolation weights are computed outside the model and read from a file
  - The interpolation is done in parallel with minimum source field information communicated to the individual MPI tasks
    - If destination points *a* and *b* on task *N* both needing source point *c* from task *M* then it is only sent to task *N* once
- The interpolation weights are read from a SCRIP convention netCDF file
  - Different tools (including existing coupling infrastructures) can be used to create these weights
    - No need to reinvent computing for weight

Current ECMWF coupled system 2:

- Coupling is sequential with the following calling sequence:
  - Atmosphere → Waves → Ocean
  - Time stepping in the atmosphere model (IFS) drives the coupling
- All model runs on the same MPI tasks / OpenMP threads configuration
  - All model needs to support this programming model and ideally do it well. Not quick the case for NEMO but we are getting there

Coupled model setup used for our INCITE runs:

- 4 or 5 start dates per storm with 7 days run length
  - Medium range like NWP.
  - Target for operations is around 240 model days/day
- ECMWF ATOS: 4 Atos BullSequana XH2000 clusters.
  - AMD EPYC Rome CPUs
  - 1920 nodes/cluster, 128 cores/node
  - Used for TCo1279, TCo2559 and TCo3999 runs with eORCA025/eORCA12 for NEMO
  - Largest job within INCITE (TCo3999_eORCA12) took about 3 hours on 500 nodes
- DOE summit:
  - 4,608 nodes with 2 IBM POWER9 CPUs (42 CPUs cores/node) and 6 NVIDIA Volta GPUs each
  - GPUs were used for the spectral transforms within the IFS only
    - More GPU code coverage is underway
  - Use for TCo7999+eORCA12 only
  - Jobs used about 960 nodes for about 18 hours

Grids used for the INCITE runs

Atm grid	Atm points	Atm res (km)	WAM points	WAM res (deg)
TCo1279	6599680	9	1873911	0.125
TCo2559	26306560	4.4	11710102	0.05
TCo3999	64144000	2.2	11710102	0.05
TCo7999	256288000	1.4	11710102	0.05

NEMO grid	NEMO points	NEMO res (deg)
eORCA025	1740494	0.25
eORCA12	15585132	1/12

Scientific challenges

- We wanted to use the 1/12 degree ocean model
  - A preliminary NEMO V4 setup was used for these runs for all resolutions
  - Ocean initial conditions were not available
    - A forced ocean run with SST relaxation was used to produce all ocean initial conditions
    - No data assimilation, but all runs consistent in the way the initial conditions were prepared
    - Work is ongoing to allows us to prepare initial conditions for any NEMO resolution based on an ocean reanalysis
  - It would have been cool to run with eORCA36 (1/36 degree), but we don't have the in-house knowledge to do so
    - Preparing ocean initial conditions would have been very time consuming
- The non-hydrostatic version of the IFS had some problems so only a limited set of NH runs were done
- The IFS deep convection scheme was active for most runs, but with reduced effect

Technical challenges

- Get the model to run on Summit:
  - Getting enough GPU coverage to make efficient use of Summit
- Dealing with the interpolation weights
  - The netCDF files are around 37 GBytes for the coupled TCo7999+eORCA12 system
    - Originally, we had to read them on every MPI tasks to compute the communication patterns for coupling
      - Very high memory watermarkin this phase of the model
    - Start up of the model was expensive, but a solution was implemented to store the communication patterns, so we only have to read them for one run with subsequent runs being able to reuse the information from the first run as long as the number of MPI tasks do not change
    - An offline tool to compute the communication pattern has been developed and used for the coupled tco7999 runs on summit
- I/O was challenging in many ways:
  - What works well on GPFS on Summit does not always work well on Lustre on ATOS
    - Examples: Many readers are typically fine on GPFS, but not on Lustre
  - Moving and processing of model output from these run is very time consuming

Cost of the ocean model (including the coupling):

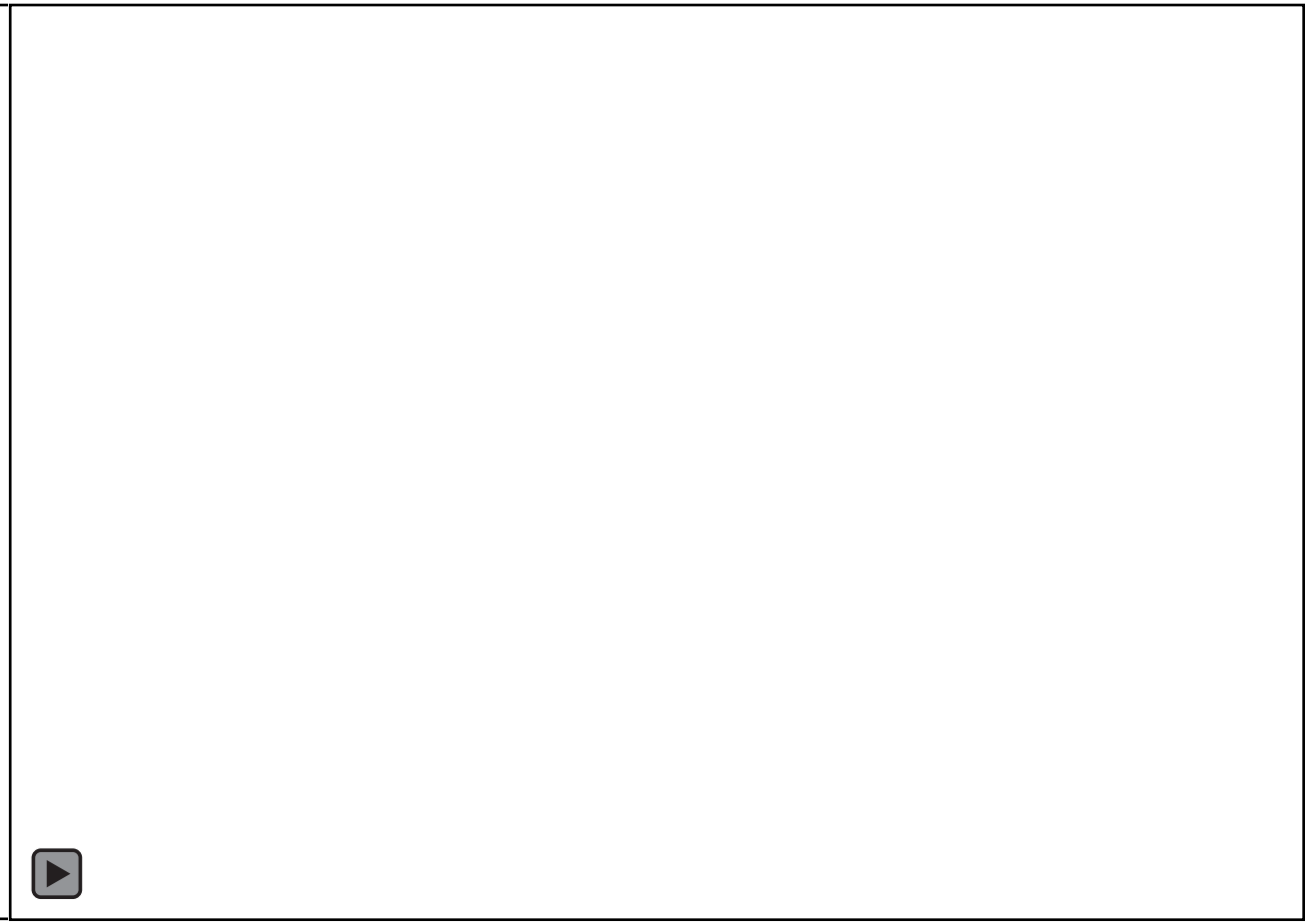
Atmosphere	Ocean	Cost of ocean (%)	HPC system
TCo1279	eORCA025	5	ATOS
TCo1279	eORCA12	46	ATOS
TCo2559	eORCA025	5	ATOS
TCo2559	eORCA12	15	ATOS
TCo3999	eORCA025	3.5	ATOS
TCo3999	eORCA12	7	ATOS
TCo3999 NH	eORCA12	2.4	ATOS
TCo7999	eORCA12	1.1	Summit

- Initialization of NEMO and the coupling is not included in this

Summary of science results (detailed analysis is in progress)

- Dealing with the model output is a pain at very high resolutions
  - Move more diagnostics into the model to limit the needed output
    - The observation operators in NEMO meant that we can compare observations to model at full temporal and spatial resolutions
- Uncoupled integrations becomes unphysical at high resolutions
  - Unphysical core pressures down to 850 hPa
- Higher atmospheric resolution gives stronger storms for coupled simulations
  - Lower core pressure
  - Stronger winds
- When coupling to the ocean the ocean response is more realistic with higher atmosphere resolution and for some storms also with higher ocean resolution
  - Validated against ocean observations
  - eORCA12 is not really very high ocean resolution, but this is what we were able to do

Cold wake for TC Florence 2018 from TCo1279+eORCA025 and TCo3999+eORCA12 runs:



Summary and outlook

- We have successfully managed to run global coupled simulations with very high atmospheric and high (but not very high) oceanographic resolutions
  - We can now address the science questions of the benefit of higher resolution for the NWP problem
- Limited changes to our coupled model infrastructure were needed to do this
  - For highest resolutions reading of full interpolation weights were problematic so a preprocessing program to compute the communications for a given set of domain specifications were created
  - How does other couplers deal with large interpolation weights files?
- Lots of work having been done over the years to introduce GPU enabled parts in the IFS
  - And lots more work are still needed
- We will probably use an infrastructure where different parts of the modelling system runs on different hardware:
  - E.g., keep NEMO on CPUs for now but possible use the PsyClone tool to move it to GPUs
  - Next procurement is going to be interesting

Questions?

ECMWF