We would like to thank the anonymous referee for her/his comments and time spent to review the manuscript.

Please, find below the answers (regular letters) to the questions (cursive):

- In the atmospheric model, you mention a variety of parametrization scheme, could you be more precise about it? micro-physics scheme? convection scheme? aerosols climatology?

We chose a parametrization scheme suitable for the domain of interest (EU): (1) Micro-physics scheme: The scheme used in the COSMO-EU model (itype gscp=3 in the namelist of CCLM). This is similar to the original Kessler (1969) scheme, but including cloud ice as an additional prognostic variable (cloud ice scheme). The scheme allows for an explicit representation of ice clouds and a more complete simulation of precipitation formation in mixed phase clouds. (2) Convection scheme: Mass flux Tiedtke scheme. (3) Aerosols climatology: AeroCom Global AOD data (An AeroCom initial assessment – optical properties in aerosol component modules of global models , Atmos. Chem. Phys., 6, 1815–1834, 2006). The manuscript has been modified to have a better description of the atmospheric model.

- There is no reference to the land component of the system, is there any routing of water? what about the vegetation? change of land use throughout the 100 years?

COSMO-CLM is internally coupled to the TERRA model, that is the land and vegetation component. There is no lateral routing of water but climatologies are used for discharge. The land use applied throughout the century does not change. This information has been added to the new manuscript.

- Could you explain the motivations for the 2 step methods providing the lateral boundary conditions? What is the motivation to do a 3D relaxation of the ocean towards a simulation performed with another atmospheric forcings?

The assimilation of the 20CR reanalyses made by the global MPI model was needed since no ocean lateral boundary conditions are available from 20CR. We better motivated this in the manuscript (Sec. 2.1).

- The exchanged variables and the according coupling time-step should be listed

We have added more details in the coupling section about it. Regarding the variables exchanged, CCLM sends to NEMO Evaporation-Precipitation, solar fluxes, net solar flux, wind stress, and only to NEMO-Nordic Surface pressure. On the other hand, NEMO sends to CCLM the SST, and only NEMO-Nordic the Ice fraction. The exchange is made every three hours.

- An important point of such a climatic run is the initial state: is there any spin-up? spin-up of the ocean only? spin-up in coupled mode?

The three marginal seas were spun up. NEMO-Nordic with a 5-year spin-up, and NEMO-Med with 20 years, both in coupled mode. This has been clarified in the new version.

- Regarding the system performances, instead of nodes could you state the number of cpu cores used? and gives some numbers about the cpu-time needed for the run.

The total number of CPUs used was 664 (CCLM=24x13, NEMO-MED= 12x8, NEMO-NORDIC=16x16). The computation of each simulated month lasted around 1h 25'. For a simulation of 110 years, this implies around 78 days in total to complete the simulation.

- The 3 models are running simultaneously and eventually wait for one another, so to be faster it would be interesting to optimize the slowest model i.e. NEMO-BALTIC (or NEMO-NORDIC should be consistent all along the manuscript). It looks like you could move some cores from NEMO-MED to NEMO-NORDIC to make it faster and at least decrease the waiting time in COSMO-CLM. Any comments on that ? and on the scalability of the system ?

Will et al. 2017 show that the optimized computational performance of the coupled system is weakly dependent on the computing architecture or on the individual model components but strongly depends on the coupling method. It is true that some NEMO-MED cores could be transferred to NEMO-NORDIC, but one has to pay attention to the number of cores that each node has, because the interconnexion of nodes is slower (and so it penalizes more) than the connection of cores within the same node. Therefore, it is optimal to choose per model a number of cores that is a multiple of the number of cores per node, like we did.

- I find surprising the 3.6 ratio between the atmosphere only and coupled since atmosphere is running on the same number of cores and there is not that much waiting time. How do you deal with the I/O ? Could it be some latency due to the writing of the other models outputs ?

The I/O is done in parallel and effectively. Adding the marginal seas implies 523x619x56 (=18.129.272) grid points for the Baltic and 264x567x75 (=11.226.600) grid points for the Mediterranean, compared to the 226x232x40 (=2.097.280) grid points of CCLM. The NEMO models have a lower output frequency than CCLM, since NEMO writes output only every 5 days.

- In the analysis of the SST, a comparison is done between the result of the RCSM simulation and the prescribed SST coming from a global simulation at low resolution. In my opinion, the comparison not only shows the impact of retroaction through coupling of atmosphere and ocean models but also the impact of using different models and resolution. Should we not compare the SST from an ocean-only simulation forced by a COSMO-CLM run (driven by a prescribed SST) and the SST from the RCSM? Could you state something about this point?

We agree that the prescribed SST coming from the global simulation has lower resolution. To know more about the impact only of the coupling (and not affected by the resolution), we would like to compare the SST coming from an ocean-only NEMO simulation and the SST of our coupled system. However, there is no available ocean-only simulation with NEMO covering the whole century. Nevertheless, we might have access to an ocean-only NEMO simulation over the North and Baltic Seas that run for a couple of decades, so we will start with a comparison there.

- Do you have any hypothesis regarding the various bias we observe?

The bias in the SST in the Mediterranean is seasonal dependent (colder in winter, warmer in summer), and many different factors could affect it (ocean initialization, higher mixing layer depth, aerosols blocking radiation, internal NEMO dynamics, etc.). We have analyzed the mixing layer depth looking for some answers, but did not succeed. To understand better the coupled system, we have run sensitivity tests changing the SST. Those results are summarized in a recently accepted publication (Kelemen et al. 2019). Nevertheless, we will continue looking for answers that explain those biases.

- about density histograms, at first it was not obvious to me that the dark pink was in fact the blue lying behind the light pink

This has been better explained in the new version.

- Could you mention the convention used for the boxplots ?

We represent the box-and-whisker plot computed with R using the default method. The box is defined by the 25th, 50th (median) and 75th percentiles. The whiskers are the lowest value still within the 1.5 IQR of the lower quartile, and the highest the value still within the 1.5 IQR of the upper quartile. The points are outliers.

- For figures of differences, could you precise the sign of the difference (simulation - observation, for instance):

Simulation-Observation referred to simulation minus observations, and differences between coupled and uncoupled referred to Coupled minus uncoupled. Nevertheless, we are more precise in the new version.

- I find interesting the analysis of extreme events, but you could perhaps show less figures or indices and make same assumptions explaining the better behaviour of the coupled model, and why on some indices and not on others?

There are 27 recommended indices, however, we had to choose some of them to avoid a very long paper. We considered monthly max and min values of the daily max and min temperature, as well as dry/wet spells, since those are relevant and describe well extreme events showing some impact of memory in the climate system and additionally having an impact on human lives. We have presented these indices for some German stations were the complete century of data was available. However, explaining the mechanism affected by the coupling that lead to changes in the local max and min temperature in those stations is not trivial, that is why the paper just focus on a description of how the indices behave, rather than in the explanation why this happens like this.