Optimization of weather forecasting for cloud cover over the European domain using the meteorological component of the Ensemble for Stochastic Integration of Atmospheric Simulations version 1.0

Yen-Sen Lu[1,4], Garrett Good[2], and Hendrik Elbern[3]

[1]Institute of Energy and Climate Research – Troposphere (IEK-8), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany
[2]Fraunhofer Institute for Energy Economics and Energy System Technology IEE, Kassel, Germany
[3]Rhenish Institute for Environmental Research at the University of Cologne, Cologne, Germany
[4]Jülich Supercomputing Centre, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

Correspondence: Yen-Sen Lu (ye.lu@fz-juelich.de)

Abstract.

~~In this study, we present an expansive sensitivity analysis of physics configurations~~ We present the largest sensitivity study to date for cloud cover using the Weather Forecasting and Research Model (WRF V3.7.1) on the European domain. The experiments utilize the meteorological part of a large ensemble framework~~known as the~~ , ESIAS-met (Ensemble for Stochastic Integration of Atmospheric Simulations~~(ESIAS-met). The experiments first seek the best deterministic WRF physics configuration by simulating~~). This work demonstrates the capability and performance of ESIAS for large ensemble simulations and sensitivity analysis. The study takes an iterative approach by first comparing over 1,000 combinations of microphysics, cumulus parameterization, planetary boundary layer physics (PBL), surface layer physics, radiation scheme, and land surface models ~~. The results on six different test days~~ on six test cases. We then perform more detailed studies on the long-term and 32-member ensemble forecasting performance of select combinations. The results are compared to CMSAF satellite images from EUMETSAT. ~~We then selectively conduct stochastic simulations to assess the best choice for ensemble forecasts.~~ The results indicate a high ~~variability in terms of physics and parameterization~~sensitivity of clouds to the chosen physics configuration. The combination of Goddard, WSM6, or CAM5.1 microphysics with MYNN3 or ACM2 PBL ~~exhibited the best performance~~ performed best for simulating cloud cover in Europe. For ensemble-based probabilistic simulations, the ~~combination~~ combinations of WSM6 and SBU–YL microphysics with MYNN2 and MYNN3 ~~showed the bestperformance, capturing the cloud fraction and its percentiles with 32 ensemble members. This work also demonstrates the capability and performance of ESIAS-met for large ensemble simulations and sensitivity analysis~~performed best.

1 Introduction

~~Recent events in 2021~~ The 2020s have begun with increasingly frequent and extreme weather in Eurasia, with a series of floods, heat waves, and droughts from Europe to China, most recently culminating in the deadly situation in Pakistan. Such events have highlighted the ~~destructive potential~~ destructiveness of extreme weather to human life and ~~property. The prevention of such loss has demonstrated the urgent need for better~~ infrastructure and the potential for weather forecasting and research to help better understand these weather conditions ~~, especially with respect to how they are caused by climate change (Tabari, 2020; Palmer and Hardaker, 2011; Bauer et al., 2015; Sillmann et al., 2017; Samaniego et al., 2018; Bellprat et al., 2019; Bauer e~~ ~~.~~ and how climate change may catalyze future tragedies (Tabari, 2020; Palmer and Hardaker, 2011; Bauer et al., 2015; Sillmann et al., 2017 ~~.~~

At the same time, the Russo-Ukrainian War has created an urgent desire across Europe for energy security and independence towards local, renewable generation. In the energy sector, better weather predictions ~~can help prevent damage to power infrastructure, but also to combat climate change in the first place by facilitating the~~ help facilitate the economical integration of higher proportions of ~~weather-dependent renewables into the power system~~ wind and photovoltaics into power systems (e.g. Rohrig et al., 2019; Adeh et al., 2019), for which unexpected weather can create bottlenecks at the small scale or be incredibly expensive ~~for grid operators and result in , for example, negative wind energy prices . Moreover, better forecasting is also needed to study to~~ in energy markets at the large scale, even resulting e.g. in negative prices during high wind events. Aside from the economics, evaluating the impact of energy on ecology ~~Yan et al. (2018); Lu et al. (2021)~~will also require better forecasting (Yan et al., 2018; Lu et al., 2021).

This study has resulted as part of a larger effort to advance towards exascale computing in weather forecasting, with a focus on energy meteorology for solar power predictions. In this context, we have aimed to optimize the performance of the ultra-large ensemble system ESIAS[1] and its meteorological component WRF[2] for cloud cover as compared to satellite measurements over the central European domain, demonstrating the framework and creating to our knowledge the largest sensitivity study of WRF to date.

~~There are generally two types of weather predictions: deterministic and probabilistic simulations (Palmer, 2012). Deterministic simulations rely on the accuracy of a single simulation to capture the values of meteorological variables~~

1.1 Sensitivity analyses for deterministic and ensemble simulation

Sensitivity analysis is a widely accepted method for identifying the most suitable model composition. Improving the accuracy of ~~such~~ weather and climate models involves research to improve numerical solvers, ~~accurate parameters~~parameter accuracy, and advanced governing equations. Various global and regional deterministic weather models are developed by national and international weather agencies. ~~However, the~~ The optimal implementation of any model can however vary greatly for different regions. The widely used ~~Weather Research and Forecast (WRF) model, a publically available~~ and publicly available WRF

[1] Ensemble for Stochastic Integration of Atmospheric Simulations (ESIAS)
[2] Weather Research and Forecasting Model (WRF)

research software system, ~~is,~~ for example, is developed in North America, where the optimal ~~meteorological models, or even the optimal~~ model configuration or even parameterization of land types ~~given~~ (e.g. the typical density and size of structures) ~~,~~ can differ in Europe and elsewhere. It is for example the authors' experience that WRF typically results in biased predictions of the solar resource in Germany.

~~Sensitivity analysis is a widely accepted method for identifying the most suitable model composition.~~ The improvement of high-performance computation has enhanced the ability to perform not only higher resolution simulations, ~~or larger simulations for sensitivity analysis~~ but also larger sensitivity analyses (Borge et al., 2008; Jin et al., 2010; Santos-Alamillos et al., 2013; García-Díez et al., 2013; Mooney et al., 2013; Warrach-Sagi et al., 2013; Kleczek et al., 2014; Pieri et al., 2015; Stergiou et al., 2017; Gbode et al., 2019; Tomaszewski and Lundquist, 2020; Varga and Breuer, 2020). To date, most sensitivity analyses are based on a small number of combinations of physics configurations. The largest sensitivity analysis of WRF, for example, includes 63 physics combinations (Stergiou et al., 2017), whereas WRF has over 1 million possible combinations of 23 micro-physics, 14 cumulus settings, 13 planetary ~~boudary~~ boundary layer (PBL) physics, 7 land surface and 8 surface layer models, and 8 ~~longwave~~ long- and 8 shortwave radiation schemes (Skamarock et al., 2008). ~~However, it is not enough to optimize WRF by sensitivity analysis on a few configuration sets. Most~~ There is thus potential for optimization, as most physics combinations can be expected to be biased ~~to the measurement from ground observations, making them unsuitable for deterministic forecasts that should perform well on average~~as compared to observations.

~~Probabilistic simulations are meant to address~~ The model optimization is typically an effort in deterministic accuracy, but there are two general types of weather simulation, deterministic and probabilistic (Palmer, 2012). Whereas the quality of a single deterministic prediction relies on substantial work to obtain accurate model data and physics, ensemble-based probabilistic simulations focus on the spread of possible solutions~~, accounting for uncertainty from~~ . This accounts for the uncertainty from the initial conditions or ~~modelling physics by performing large multiphysics ensemble simulations~~ model physics using large or even multiphysics ensembles of solutions (e.g. Li et al. (2019)), or employing stochastic schemes ~~in light of greater high-performance~~ if afforded higher performance computing power (Ehrendorfer, 1997; Palmer, 2000; Dai et al., 2001; Gneiting and Raftery, 2005; Leutbecher and Palmer, 2008; Hamill et al., 2013). While deterministic forecasts should provide the most likely case, probabilistic forecasts ~~are likely~~ aim to capture the uncertainty of that solution~~, including rare weather extremes. The challenge of ensemble simulation is thus not only~~ . The optimal model configuration may then differ for the ensemble application, depending e.g. on the model physics variance or sensitivity to perturbation.

1.2 Technical challenge

Finally, besides the scientific challenge ~~(e.g. proper scoring rules, Sillmann et al., 2017) but~~ of ensemble forecasting (or e.g. assessing prob
, there is also the technical challenge ~~, for example large supercomputing facilities that can handle~~ of e.g. powerful supercomputing facilities and storing simulations large enough to capture the ~~most~~ detail and outliers needed to detect extreme and damaging events typically missed by contemporary, $O(10)$ member ~~simulations.~~

~~The urgent need for better probabilistic simulations is the motivation behind the development of the Ensemble for Stochastic Integration of Atmospheric Simulations (ESIAS)(Berndt, 2018; Franke et al., 2022). We have developed ESIAS to carry out~~

ensembles. Presently, large supercomputers can computationally produce ultra-large ensembles of $O(1,000)$ members (at moderate resolution), if challenges in I/O performance and MPI communication are addressed. The ESIAS framework (Berndt, 2018; Frank has been developed to accomplish ultra-large ensemble forecasts of ~~(O)1000 members and have further integrated stochastic schemes to generate simulation membersfor data assimilation, both for probabilistic simulations with stochastic schemes and for sensitivity analysis~~up to (O)1000 members, demonstrated in this study with both multi-physics and stochastic schemes for probabilistic simulation of cloud cover. Moreover, ESIAS aims to ~~cope with~~ meet future *exascale* computation requirements in order to perform forecasts that are not yet operationally possible. ~~We use ESIAS to perform a sensitivity analysis of over 1,000 WRF physics combinations in Europe, with a focus on wind and solar energy and to provide recommendations for future weather research.~~

~~The object of this study is to optimize ESIAS-met by determining the most suitable physics configuration to better perform the simulation of the cloud cover by comparing simulation results to the satellite measurements. We also perform simulations combining multi-physics simulation and stochastic simulation to achieve the probabilistic simulation to capture the cloud cover condition over the European domain.~~

1.3 Outline

In this article ~~, we introduce the ensemble weather forecasting system ESIAS and its meteorological component~~ we proceed in section 2 ~~. This section describes the methods applied for the sensitivity analysis regarding~~ with descriptions of the forecasting system, the model physics ~~configuration and the description of the methods for evaluating the simulation results~~configurations, and the methodology used for the sensitivity analysis. Section 3 describes the data used in this study. The ~~outcome of the three sets of sensitivity analyses is shown~~ sensitivity analysis itself is performed iteratively in four sets in section 4~~and the~~ , beginning with a general test of a very large assortment of models before winnowing this down with increasing detail. The final results are discussed and concluded in ~~the~~ section 5.

2 Model description

2.1 ~~Modeling~~ Modelling system: ESIAS-met v1.0

ESIAS is ~~the~~ a stochastic simulation platform developed by IEK-8 ~~at Forschungszentrum~~ of the Jülich Research Centre and by the Rhenisch Institute for Environmental Research at the University of Cologne. ESIAS ~~consists of~~ includes two parts, which are based on the Weather Research and Forecasting (WRF) model V3.7.1 (Skamarock et al., 2008) and the EURopean Air pollution ~~Dispersion~~ and Dispersion – Inverse Model (EURAD–IM, Franke et al. (2022)), which we shall refer to as ESIAS-met and EISAS-chem, respectively. The full details of these two models are described by Berndt (2018) and Franke (2018).

Figure 1 illustrates the ~~full concept~~ workflow of ESIAS-met, based on WRF V3.7.1. The ESIAS System Control Scripts are typically used to control the WRF Preprocessing System to produce the intermediate meteorological inputs ~~to generate WRF~~

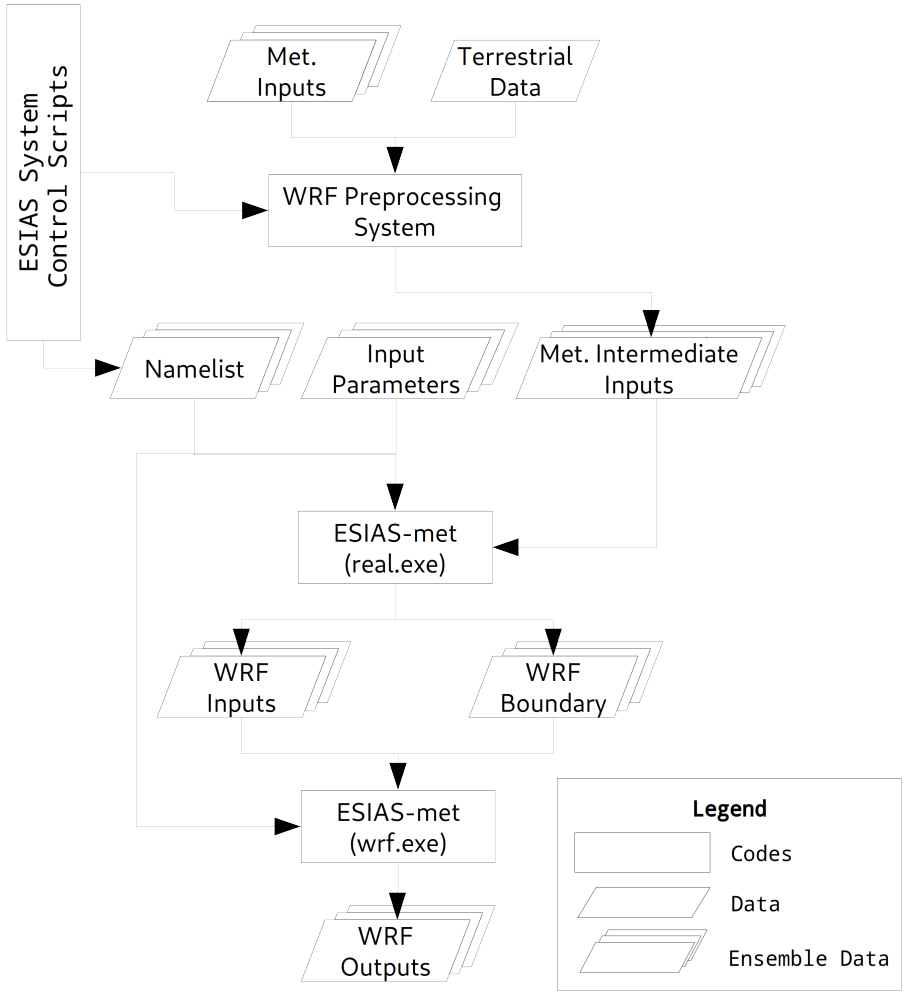

Figure 1. The concept of ESIAS-met scheme and ~~entire~~ the whole process of ~~the~~ ESIAS-met~~scheme~~. The ESIAS System Control Scripts control the WRF ~~Preprocessing~~ Processing System and the ~~namelist~~ Namelist to generate necessary inputs.

~~inputs~~ and boundary data for the simulation, though for the model setup described in the next section, both GEFS and ECMWF ERA5 data can be used as the input. The namelist generated by the `WRF_TOOLS` of ESIAS-met is the same as for WRF V3.7.1 and ~~thus~~ the input and output filenames are flexible for ~~different ensemble simulation strategies. The ESIAS System Control Scripts includes a namelist generator based on the ecosystem of ESIAS-met to better fit the system.~~ outputting large numbers of files.

The ESIAS-met executables apply the Message Passing Interface (MPI) ~~to perform large ensemble simulations and are thus advantageous~~ for large ensemble simulations with interactive members on HPCs ~~.~~ (Large individual ensemble simulations on ~~the~~ HPCs ~~is inhibitive and will~~ are inhibitive and cause long queuing times). The main purpose of ESIAS-met is to

perform large ensemble simulations ~~based on stochastic schemes. The stochastically perturbed parameterization tendency~~ with stochastic schemes, and thus both the Stochastically Perturbed Parameterization Tendency (SPPT) ~~scheme~~ (Buizza et al., 1999) and the ~~stochastic kinetic energy backscatter scheme~~ Stochastic Kinetic Energy Backscatter Scheme (SKEBS) (Berner et al., 2009, 2011) are ~~therefore used.~~ implemented.

2.2 Model setup

ESIAS-met is ~~run~~ driven with boundary conditions and intermediate meteorological inputs from the Global Ensemble Forecasting System (GEFS) and ~~the~~ with MODIS land use data~~for generating meteorological intermediate inputs~~. The map projection is Lambert ~~Confromal with a central point of~~ Conformal with central point ($54^oN, 8.5^oW$). The horizontal resolution is 20 km ~~,~~ and the number of horizontal gridpoints ~~are 180 by 180. The vertical layersconsist of 50 grid points, which are unevenly spaced for the first 11 layers~~ is 180×180. There are 50 vertical layers, only the first eleven of which have uneven spacing near the boundary layer. ~~We~~ In this study, we do not use ~~the nested domain to generate finer grids in higher resolution due to the high~~ any finer, nested domain due to computational demand. We thus could not evaluate Eta or Morrison microphysics, as the former requires higher resolutions ($< 5km$) and the latter cloud-resolving simulations (WRF, 2015). A previous study and large sensitivity analysis by Stergiou et al. (2017) ~~, which~~ tested 68 ~~cases of~~ different physics configurations ~~, performed the simulation~~ also on the European domain~~but~~ , though with a different approach~~vis-á-vis the WRF physics configuration.~~ , changing the physics options one at a time.

~~We generate a very large ensemble simulation using different physics schemes to~~ Large ensemble simulations with members of different physics were created in three sets to iteratively investigate the optimal ~~physics configuration for the simulation output. We generate three sets of large ensemble simulations based on different combinations of physics schemes~~configuration for cloud cover and the PV forecasting application. The first set (Set 1) ~~investigates~~ is the broadest with 560 combinations of microphysics, ~~the cumulus parameterizations~~cumulus parameterization, and planetary boundary layer physics. ~~The physics configurations used~~ Accordingly, only a few test cases with differing but typical cloud conditions could be afforded for Sets 1-3, whereas both the seasonality and probabilistic performance were tested for the final four configurations in Set 4. The Set 1 combinations and the acronyms for the WRF physics and parameterizations are listed in Table 1. ~~It is recommended that the~~ We note that the official documentation recommends setting the surface layer physics ~~be set with~~ with specific planetary boundary layer physics in WRF ~~. The five-layer thermal diffusion~~ (WRF, 2015). Five-layer thermal diffusion is employed for land surface physics, Dudhia for shortwave radiation physics, and RRTM for longwave radiation physics ~~are employed~~ in this set of numerical experiments.

Set 2 takes a subselection from the Set 1 ~~comprises a total of 672 ensemble simulations for the optimization.~~

~~The second set (Set 2) adds combinations of~~ results and adds land surface models, shortwave radiation schemes, and longwave radiation schemes to ~~a combination of microphysics, cumulus parameterizations, and planetary boundary layer physics~~

~~selected based on the results from Set 1 (Table 2). In Set 2 there are~~ form the additional 513 ~~ensemble simulations in total. For the acronyms of WRF physics and paratermizations, view Table ??.~~

~~The~~ combinations described in Table 2. We note here that the PBL ACM2 only considers the surface layer physics of MM5 similarity and ~~therefore~~ thus does not apply to the other surface layer physics models. The ~~combination of Eta similarity and the CLM4~~ land surface model ~~is ruled out as it~~ CLM4 also does not employ ~~surface layer physics Eta similarity~~ Eta similarity and thus we exclude this combination. For the ~~longwave and shortwave~~ long- and short-wave radiation physics, we have only three combinations~~of physics configurations~~: RRTM and Dudhia, RRTMG and RRTMG, and Goddard and Goddard for ~~shortwave~~ the short- and longwave radiation physics ~~scheme~~schemes, respectively. For surface layer physics, we employ ~~Monin–Obukhov similarity and therefore the revised~~ Monin-Obukhov (MO) similarity and hence the Revised MM5 MO similarity (listed as MM5 similarity~~and Eta similarity schemes~~) and Janjic-Eta MO similarity are utilized. The MYNN surface layer scheme is used to investigate its suitability with the MYNN2 and MYNN3 PBL physics. The sophisticated land surface models CLM (version 4) and Noah LSM are tested ~~as well as~~ along with RUC LSM, which performs similarly to the other two LSM (Jin et al., 2010).

Table 1. Employed physics configuration of Set 1, which summarized ~~$12 \times 7 \times 8$~~ $10 \times 7 \times 8$ configurations. The ~~full description of abbreviations can be found in Table ??~~abbreviation is follow by the physics name.

Clusters	Microphysics	Abbr.	Cumulus Parameterization	Abbr.	PBL	Abbr.	&	Surface layer
	Kessler	Ke	Kain-Fritsch	KF	YSU	YSU		MM5 similar
	Lin (Purdue)	Lin	Betts-Miller-Janjic	BM	MYJ	MY	~~Eta~~	Janjic-Eta sin
	WSM3	W3	Grell-Freitas	GF	GFS	G		Pleim-Xiu
Physics	WSM5	W5	Simplied Arakawa-Schubert	OS	QNSE	Q		QNSE surfac
~~Eta (Ferrier)~~ options	WSM6	W6	Grell-3	G3	MYNN2	MN2		MM5 similar
~~WSM6~~	Goddard	Go	Tiedtke	T	MYNN3	MN3		MYNN surfa
~~Goddard~~	Thompson	Th	New SAS	NS	ACM2	A2		MM5 similar
~~Thompson~~	Milbrandt 2-mom	Mi			BouLac	BL		MM5 similar
~~Milbrandt 2-mom~~	CAM 5.1	Ca		~~Morrison 2-mom~~				
~~CAM 5.1~~	SBU–YLin	SB		~~SBU–YLin~~				

~~The third set (Set~~ Following the Set 2 results, six microphysics and six combinations of PBL and surface layer physics were chosen for 36 further combinations for stochastic study in Set 3~~) is used to understand the impact of the physics configurations on the ensemble simulations. We also perform a small number of ensemble simulation with a size of $O(32)$ by SKEBS for each physical configuration,which is based on the combination of different microphysics and PBL physicslisted in Table 3~~, listed in Table 3. Here, 32-member ensembles were created with SKEBS for a total of 1,152 members. All Set 3 simulations employ the Grell-3 cumulus parameterization, the Dudhia shortwave radiation physics, the RRTM longwave radiation physics, and the RUC land surface model for land surface physics. Although ESIAS-met can employ both SPPT and SKEBS schemes,

Table 2. Employed physics configuration of Set 2,which summarized $3 \times 3 \times 3 \times 19$ configurations. ~~The full description of abbreviations can be found in Table ??~~ Both Revised MM5 MO and ~~Table ??~~Janjic-Eta MO similarity is based on Monin-Obukhov similarity and therefore we use MO to represent the abbreviation.

Clusters	Microphysics	Cumulus	PBL	Radiation(SW&LW)	Abbr.	Surface Layer	~~LSM~~ Abbr.	LSM	Abbr.
Physics	WSM5	Kain-Fritsch	MYNN2	RRTM+Dudhia	RD	MM5 similarity	MMO	Noah	NOA
options	WSM6	Grell-3D	MYNN3	RRTMG	~~Eta~~ RR	Janjic-Eta similarity	JMO	RUC	RUC
	Goddard	Tiedtke	ACM2	New Goddard	GG	MYNN	MYN	CLM4	CL4

we only apply SKEBS in this experiment. According to Jankov et al. (2017) and Li et al. (2019), SKEBS can produce a large ensemble spread~~and we only~~ . Berndt (2018) also reports that SKEBS can more effectively produce instability in ESIAS-met than SPPT. We therefore use it to ~~identify~~ study the extent of the spread that one single stochastic scheme can produce. One ensemble member of the 32 is not perturbed ~~in order to serve~~ as to double as a control run. ~~All Set 3 simulations employ the Grell-3 cumulus parameterization, the Dudhia shortwave radiation physics, the RRTM longwave radiation physics and the~~

Finally, four combinations were selected for long-term simulations as part of a project concerning energy predictions as well as quantile calibrations (Dupuy et al., 2021). This data includes day-ahead predictions for every other day in 2018 and we use it here against a half year of available satellite data to verify the reliability of the recommendations under more diverse conditions than the limited test cases. The four Set 4 combinations in Table 4 are WSM6-MYNN2, WSM6-MYNN4, Goddard-MYNN2, and Goddard-MYNN3. The cumulus parameterization and surface physics are Grell-3D and Revised-MM5 MO similarity, respectively. The RUC land surface model is used for land surface physics. These simulations were conducted over the same European domain but with 15km resolution and using ECMWF instead of GEFS as input due to the availability of the data. The model gridcells increase from 180×180 to 220×220 in horizon.

Table 3. Employed physics configuration of ~~set~~ Set 3, which summarize 6×6 configurations.

Clusters	Microphysics	PBL & Surface layer physics
	Kessler	YSU MM5 similarity
	WSM3	GFS Pleim-Xiu
~~Eta (Ferrier)~~ Physics	WSM6	MYNN2 MM5 similarity
~~WSM6~~ options	Goddard	MYNN3 MYNN surface layer
~~Goddard~~	CAM 5.1	ACM2 MM5 similarity
~~Thompson~~	SBU–YLin	BouLac MM5 similarity
~~Morrison 2-mom~~		

~~We use the term *cluster* to refer to a set of physics configurations sharing a specific option, for instance all configurations sharing a particular microphysics scheme.~~

Table 4. Four select physics configuration of Set 4 ($2 \times 2 \times 1$).

Clusters	Microphysics	PBL	Cumulus param.	&	Surface layer physics	&	LSM
~~CAM 5.1~~ heightPhysics	WSM6	MYNN2	Grell-3D		MM5 similarity	&	RUC LSM
~~SBU–YLin~~ options	Goddard	MYNN3					

2.3 Model performance evaluation and measurements

2.3.1 Ternary ~~determination of~~ cloud mask

To determine the accuracy of the cloud cover prediction, we apply the determination method and separate the cloud cover fraction ~~into three gradation of grid conditions~~in each grid cell into three gradations: clear-sky ($< 5\%$), partially cloudy ($\geq 5\%$ and $< 95\%$), and fully cloudy ($\geq 95\%$) to show more ~~details beyond clear sky and cloudy~~detail beyond a binary cloud mask. The definition of clear sky follows the ~~ASOS definition~~ definition from Automated Surface Observing System (ASOS) (Diaz et al., 2014) as 5% cloud fraction, while the full cover is defined ~~as~~ analogously at 95% cloud fraction. The ~~gradation~~ inclusion of partial clouds ~~can show more details for both~~ adds detail to the comparison of the simulation and satellite data, which ~~we then compare~~of course decreases agreement rates relative to a binary mask. Table 5 ~~shows the detection rate of cloud cover conditions and the deterministic result from~~ illustrates the ternary detection possibilities for deterministic simulations. The traditional binary detection classifies the outcome into just three categories: false (~~overpredict~~overprediction), miss, and match. ~~Our~~ Ternary determination increases this to five categories ~~to capture more detail in the~~ by including partially cloudy areas, as well as the prediction ability for the different physics configurations. Here, we use the convention that "under" and "over" represent the partial matches between fully missed or false clouds.

Table 5. The five possible outcomes for the detection of predicted cloud cover ~~determined by the observed cloud cover~~coompared to observation, where 0% - 5% , 5% - 95% , 95% - 100% are defined as clear, partial, and full cover, respectively. ~~Indaddition~~ In addition to the classifications as match, miss, false/over-predict in the detection method for binary condition, we ~~add~~ have partial matches "~~over~~under" ~~between match and over-predict~~ and "~~less~~over"~~between miss and match for the ternary condition~~.

		Prediction		
		clear	**partial**	**full cover**
	clear	Match	Over	~~Over-predict~~ False cloud
observed	**partial**	~~Less~~ Under	Match	Over
	full cover	~~Miss~~ Missed cloud	~~Less~~ Under	Match

2.3.2 Kappa score

The Kappa (κ) score is ~~first used to determine the~~ used to measure agreement between two or more raters, ~~which is used to rate the score of an object by determination ,~~ using determination in large data sets ~~such as from~~ like for subjects in psychological

research (Fleiss and Cohen, 1973). This score for deterministic results is ~~widely used~~ used widely in natural sciences such as land science for determining the change of land use (e.g. Schneider and Gil Pontius (2001), Yuan et al. (2005), and Liu et al. (2017)) or machine learning for scoring and validation (e.g. Dixon and Candade (2008) and Islam et al. (2018)). The equation of the Kappa score for multiple raters is calculated as:

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \tag{1}$$

where $\bar{P}$ is the sum of P_i, the matching rate of the ith subjects or individuals being rated, for k categories~~and~~ . $\bar{P}_e$ is the sum of the category rate p_j over j and N is the total number of subjects.

$$\bar{P} = \sum_{i=1}^{N} \frac{1}{n(n-1)} [\sum_{j=1}^{k} n_{ij}^2 - (n)] \tag{2}$$

$$p_j = \frac{1}{n(n-1)} [\sum_{j=1}^{k} n_{ij}^2 - (n)] \tag{3}$$

kappa has a maximum value of one 1 and can also be negative. The maximum kappa means a full match between two datasets. kappa between 0 and 1 indicates a partial match between the two datasets, while negative kappa indicates some anti-correlation in the matching (Pontius, 2001). A good model result should result in positive κ.

2.3.3 Kernel density estimation

~~The kernel~~ Kernel density estimation (KDE) is a method to approximate the probability density function of ~~data~~dataset. A variable X with n independent data points x_1, x_2, ..., x_n at x can be expressed as

$$f_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x_i - x}{h}), \tag{4}$$

~~,~~

where h and K are the bandwidth and kernel functions, respectively. The Kernel function $K(u)$ can be uniform ~~$\frac{1}{2}I(|u| \leq 1)$ or Gaussian $\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^2)$ for different proposes. In our study we use~~ $\left(\frac{1}{2}I(|u| \leq 1)\right)$ or Gaussian $\left(\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^2)\right)$, depending on the purpose. This study uses the Gaussian kernel. Here we also propose normalizing the KDE with the cumulative KDE with x in the range from 1 to m as ~~:~~

$$f_{h,acc}(i) = \sum_{j<i} f_h(x_j), \text{for } i = 1, 2, ..., m. \tag{5}$$

The resulting cumulative KDE can be normalized by the last item of $f_{h,acc}(i)$ i.e. $f_{h,acc}(x_m)$, and therefore a normalized cumulative KDE can be used to show the cumulative probability distribution of the data, which increases monotonically.

3 Data description

3.1 Input data

The initial and lateral boundary conditions are generated from the control data of the Global Ensemble Forecast System (GEFS) of the National Centers for Environmental Prediction (NCEP) (Hamill et al., 2013). This dataset has approximately 40 km resolution and 42 vertical levels. The detail on the GEFS data is described by Hamill et al. (2011). To better represent the forecasting skill from April to September ~~, we simulate 2015-04-13, 2015-05-15, 2015-06-17, 2015-07-19, 2015-08-23, and 2015-09-21. For each simulation we use two days with 3-hourly forecasting fields based on the reforecasting data from GEFS~~in 2015, we conduct 48-hour simulations beginning April 13, May 15, June 17, July 19, August 23, and September 21. These are more or less random days in different months without rare conditions, as we target the general forecasting performance for PV. Due to limited computational resources, we are only able to demonstrate Sets 1 & 2 & 3 in ESIAS-met for day-ahead simulations beginning on these six days. Each simulation uses inputs from the GEFS reforecast data with a three-hour resolution. The soil texture and land use condition are based on the ~~2-minute resolution data and~~ Moderate Resolution Imaging Spectroradiometer (MODIS) ~~30-second resolution for~~ and Noah-modified 20-category IGBP-MODIS land use data with two-minute and thirty-second resolutions, respectively. The target domain covers most of Europe ~~, and is a single domain~~ with a 20 km horizontal resolution. Figure 2 shows the whole area of the target domain and the elevation height. The European Center for Medium-Range Weather Forecasts (ECMWF) reanalyzed ERA5 data is also applied to the simulation conducted in 2018 for a yearly day-ahead weather forecasting simulation. We apply the 3-hourly input and lateral boundary conditions as inputs for ESIAS-met.

3.2 Satellite data

To validate and rate the model performance, we use the Cloud Fraction Cover (CFC) product from the EUMETSAT Climate Monitoring Satellite Application Facility (CMSAF) (Stengel et al., 2014). The data is corrected and generated from SEVIRI on METEOSAT-8, which uses the visible, near infrared, and infrared wavelengths to retrieve cloud information. The hourly CFC data has level 2 validation (Stöckli et al., 2017) for the accuracy of total synoptic cloud cover and the data is corrected by the algorithm from Stöckli et al. (2019) using the clear-sky background and diurnal cycle models for brightness temperature and reflectance. The calculation of CFC employes a Bayesian classifier. This product covers the spatial domain of Europe and Africa ~~and also dates of simulations in~~ back to 2015, although this product was discontinued after ~~2018-03-05.~~ March, 2018. The data are cropped to the ~~European~~ central European model domain. Since the CFC data do not include the northern part of Europe ~~, we do not use the whole setup domain for this~~ and have limitations at high viewing angles above the 60th parallel (Stöckli et al., 2017), we exclude this part of the domain in the analysis.

The ~~horizontal resolution (both 0.05° in longitude and 0.05° in latitude) of CFC data is higher~~ satellite data is provided on a regular grid with a $0.05^\circ \times 0.05^\circ$ horizontal resolution, which is finer than the simulation setup of ~~20km by 20km, which is~~ 20×20 km, equivalent to 0.31° ~~in~~ longitude and 0.18° ~~in~~ latitude at the ~~center of the model domain . For the overall model domain each~~ model ~~grids between~~ domain center. In order to compare the satellite data to the lower resolution model results,

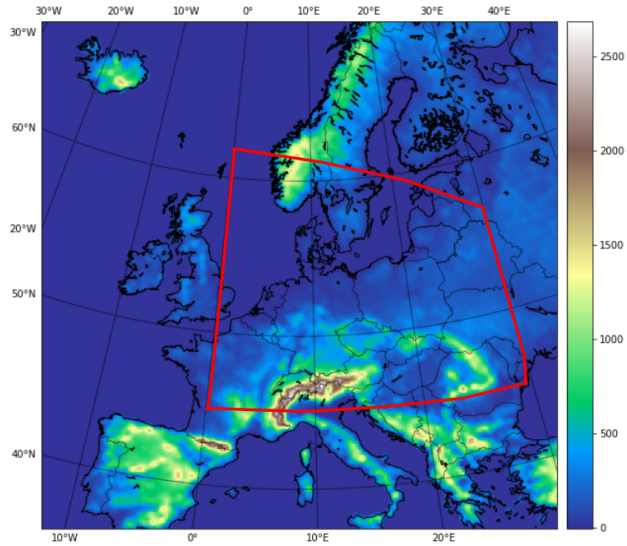

Figure 2. The topography of the target domain, ~~Europe~~ European domain, for simulation. The red line indicates the area for validating the result with cloud cover from the satellite

we simply average the CFC pixels within each grid cell[3]. The target value for any model grid point is then averaged over 12 ~~and a maximum of~~ to 36 ~~grid pointsfrom CFC to calculate the average cloud cover fraction for comparison where 99.5% of model grids contains more than 12 grid points from CFC~~ observation points, depending on the location.

The viewing zenith angle of the satellite of course creates some uncertainty in the actual observation locations due to cloud heights. This mostly vertical shift can be up to a few pixels for high clouds within the EUMETSAT grid itself, however it is at most one twenty-kilometer model point in uncertainty, such that this has only a small effect on the discrete cloud mask of the aggregated value and is negligible for the kappa scores, considering the actual resolution of cloud details arising out of the model. Should higher resolution simulations, however, be investigated in the future, this will have to be taken into account using e.g. a satellite simulator on the model data.

Figure 3 shows the ~~cloud cover condition of all simulation cases for 48 hours~~ cumulative domain cloud cover over time for each test day in 2015 ~~within the simulation domain~~according to the observation. The blue and orange curves represent the cloud cover during daytime, ~~and~~ while the red and cyan curves represent the cloud cover at night. For the cumulative plots of KDE, which is normalized to 1.0, curves with a higher accumulative rate (rapid ~~growths in~~ growth on the y-axis) represent high clear sky rates in the data~~, and the curves~~. Curves with a lower accumulative rate (y-axis) ~~depict high~~ represent more full cloud cover. ~~In the six simulation cases, the cases 2015-06-17 and 2015-08-23~~ The June 17 and August 23 cases exhibit a very high variation ~~during~~ over the 48 hours of simulation time. Moreover, ~~2015-06-17~~ June 17 shows a higher cloud cover than ~~2015-08-23. Cases 2015-04-13, 2015-05-15, 2015-07-19 and 2015-09-21~~ August 23. The April 13, May 15, July 19, and September 21 cases are similar in the cloud cover condition, since the cumulative distributions show that the variation of cloud

[3]Other studies like (Bentley et al., 1977) may use all points within a fixed radius, which may or may not overlap

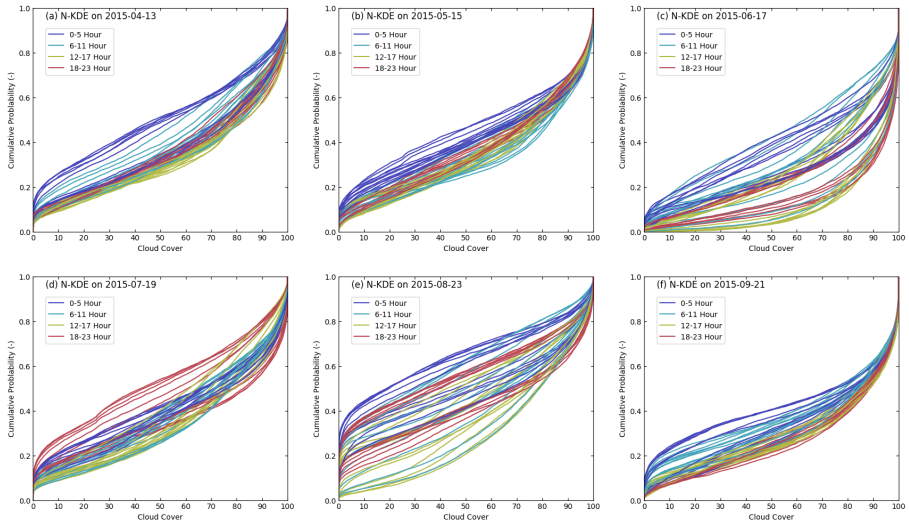

Figure 3. The cumulative KDE of cloud cover for the six simulation cases in the UTC time. The colors distinguish the different cloud cover. The different color represent different time period in a day, as shown in the legend.

cover is smaller than ~~cases 2015-06-17 and 2015-08-23.~~ for June 17 and August 23. However, both ~~case 2015-04-13 and case 2015-07-19~~ the April 13 and July 19 cases show a higher cloud cover in the early morning and in the early evening than the May 15 and September 21 cases~~2015-05-15 and 2015-09-21~~, respectively. In general, the August 23 and September 21 cases ~~2015-08-23 and 2015-09-21~~ can represent the cloud cover condition with high variability and low variability, respectively.

Finally, for the year-long simulation we used EUMETSAT cloud mask data that was available for comparison to the simulation results. This binary cloud mask is also an operational product from SEVERI that cover the full disk and distinguishes cloudy and cloud-free pixels as derived by the Meteorological Product Extraction Facility.

4 Results

4.1 Simulation efficiency

~~We perform the simulations on JUWELS (Jülich Supercomputing Centre, 2019), the~~ The simulations were performed on the JUWELS (Jülich Supercomputing Centre, 2019) high–performance computer ~~which utilizes~~ utilizing Intel Xeon 24-core Skylake CPUs (48 cores per node) and 96 GiB of main memory. We use 12 CPUs per ensemble member, ~~8,064~~ meaning 6,720 total CPUs for ~~672~~ 560 ensemble members and ~~9,216 CPUs for 768~~ 6,156 CPUs for 513 ensemble members to perform the large ensemble simulation for ~~each of the sensitivity analyses.~~ Sets 1 & 2. These large ensemble simulations are ~~successfully conducted~~ completed by JUWELS without performing farming on the HPC, ~~while~~ with the stability of ~~large ensemble simulations is~~ the large simulations guaranteed by ESIAS.

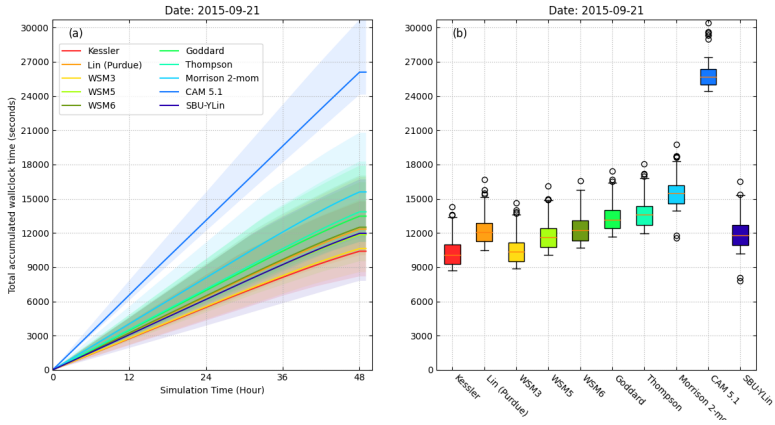

Figure 4. The (a) ~~hourly simulation time, (b)~~ total accumulated simulation time, and (~~e~~b) boxplot by different microphysics configuration within each simulation hour on ~~2015-09-21.~~ September 21, 2015. The upper and lower boundaries of the color fill indicates the maximum and minimum simulation time by other physics configurations.

Different physics configurations not only affect the resulting weather ~~fields and states,~~ data but also significantly impact the ~~computing~~ computation time. Figure 4 and Figure 5 show the average time consumption (solid line) and the range of the time ~~consumption~~ consumptions (color fill) for the configurations of microphysics and planetary boundary layer physics, respectively. We use the simulation case ~~in 2015-09-21 as the examplecase only~~on September 21 as an example. The most time-consuming ~~simulation is always that with the one~~ simulations always include CAM5.1 (average of 26,095 seconds) microphysics and the quickest ~~is with~~ have Kessler (average of 10,404 seconds), which only parameterizes the autoconvection, precipitation ~~cloud~~clouds, the evaporation of precipitation, and the condensation-evaporation function in the continuity equation. For the simulation of planetary boundary layer physics, the slowest configuration is QNSE (average of 16,967 seconds) and the fastest is GFS (average of 12,440 seconds). The cumulus parameterization has the smallest effect on time consumption, with the most time-consuming being Grell-3 (average of 14,715 seconds) and the least time-consuming being Betts-Miller-Janjic (average of 13,447 seconds), where the difference is only 9.4%.

The differences between the first and third quartiles show how much the different physics configurations affect the simulation speed, as shown in Figure 4 (c) and Figure 5 (c). For the physics clusters of microphysics, PBL physics, and cumulus parameterization, the average quartile difference is ~~1644.45 seconds~~near 1,500, 3,~~281.25 seconds~~500, and 3,~~930.54~~ 700 seconds, respectively. The ~~outliners are from the~~ outliers involve CAM5.1 microphysics, which is the most computationally expensive~~microphysics~~. The most time-consuming ~~consideration is from~~ part is the configuration of the microphysics. The average time consumption is similar for the ~~cluster~~ clusters of PBL physics and ~~the cluster~~ of cumulus parameterization, except for the ONSE PBL physics, which consumes 5,000 seconds more simulation time than other PBL physics. The boxplot confirms that CAM5.1 microphysics and ONSE PBL physics are the most time-consuming~~physics~~.

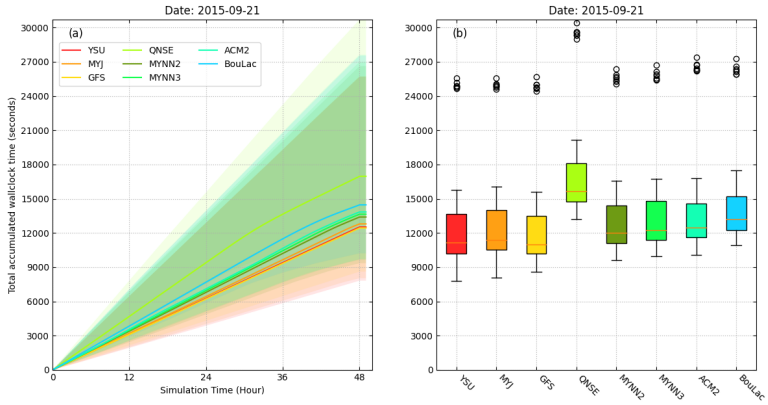

Figure 5. The (a) ~~hourly simulation time, (b)~~ total accumulated simulation time, and (~~e~~b) boxplot by different microphysics configuration within each simulation hour on ~~2015-09-21.~~ September 21, 2015. The upper and lower boundaries of the color fill indicates the maximum and minimum simulation time by other physics configurations.

4.2 ~~Sensitivity~~ Set 1 sensitivity analysis~~on the clusters~~ : Clusters of microphysics, PBL, and cumulus parameterization

The ~~672 ensemble simulations are~~ simulation results used for comparison with the satellite data exclude the first 12 hours, as the model is run in weather forecasting mode and hence can take 6-12 hours to converge (Jankov et al., 2007; Kleczek et al., 2014) . Only the last 36 hours of the simulation output are therefore used to compare with the satellite data.

The 560 simulations were performed with the target dates and times from section 3.1. The heat maps in Figure 6 and Figure 7 represent two different ~~results based on the cloud cover~~ test case results with high and low variabilities based on the cloud cover, respectively. Both figures ~~indicates the resulting Kappa by the microphysics cluster along with the cluster of~~ indicate that the microphysics Kappa scores cluster by their cumulus parameterization and PBL physics. The cloud mask results indicate that the ~~cluster of~~ CAM5.1 ~~outperforms~~ cluster outperformed the cloud cover prediction of the other microphysics, but also that Goddard and WMS3 ~~also perform~~ performed well. The ~~kappa~~ κ indicates that the Kessler microphysics predicted the worst cloud cover overall, irrespective of the PBL physics or cumulus parameterization used. The combination of the three cumulus parameterization Grell-Freitas, Grell-3, and New-SAS, and the PBL physics GFS diminishes the prediction of the cloud cover. The results from the other four cases can be found in ~~Figure~~ Figures S1-S4.

The ~~simulation results used for comparison with the satellite data exclude the first 12 hours, as the model is runs in weather forecasting mode and hence the first 6-12 hours of simulation hours can be considered to be the spin-up time (Jankov et al., 2007; Kleczek et al., 2014). Only the last 36 hours of the simulation output are therefore used to compare with the satellite data.~~

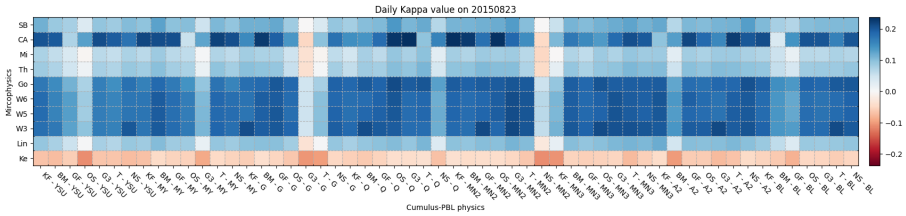

Figure 6. The heat map of average ~~kappa~~ κ for every configuration for the clusters of microphysics (~~Y-axis~~y-axis) with the cluster of cumulus parameterization and PBL (x-axis) on the simulation case ~~"2015-08-23".~~ August 23 of Set 1. The data is only use the last 36 hours of simulations for calculating the ~~kappa~~κ.

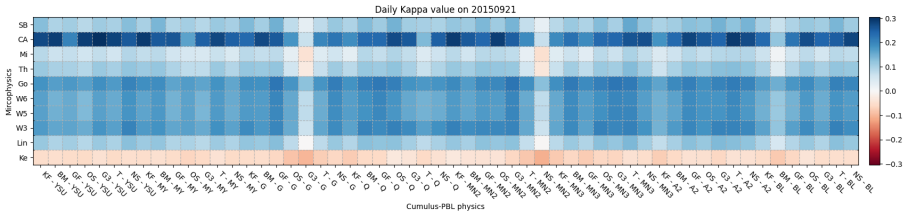

Figure 7. The heat map of average ~~kappa~~ κ for every configuration for the clusters of microphysics (~~Y-axis~~y-axis) with the cluster of cumulus parameterization and PBL (x-axis) on the simulation case ~~"2015-09-21".~~ September 21 of Set 1. The data is only use the last 36 hours of simulations for calculating the kappa.

~~The~~ overall results for ~~Kappa~~ κ from the six test cases (Figure 8a) confirm that CAM5.1 ~~performs~~ performed best for cloud cover, and that both the WRF single moments 3 (WSM3) and the Goddard microphysics also ~~perform~~ performed well. The only exception is ~~simulation case2015-04-13~~April 13 case, where the Goddard microphysics cluster ~~outperfomed~~ outperformed the other microphysics. The simulations for ~~case 2015-06-17 and 2015-08-23 are~~ the June 15 and August 23 cases were performed well by the Goddard microphysics and the WMS3/5/6 microphysics clusters~~, where~~ . Here, the cloud cover varies more than in the other cases. The performance of the Goddard microphysics and the WMS3/5/6 microphysics clusters ~~are~~ was equal to the CAM5.1 microphysics cluster ~~when the~~ in the cases with large cloud cover variation~~is large, as in cases 2015-06-17 and 2015-08-23.~~ , as for June 17 and August 23.

Figure 9 and Figure 10 show the average cloud fraction from hour 12 to 48 from both the satellite data and the simulation result for ~~2015-08-23 and 2015-09-21~~August 23 and September 21, respectively. In this simulation case, we use the different microphysics and the cumulus parameterization of Grell-3D and the PBL physics of MYNN3. In the ~~2015-08-23~~ August 23 case, the sky is partially clear above the North Sea and in Eastern Europe, and rather cloudy around the Alps. From the selected microphysics, Kessler, WSM6, Goddard, CAM5.1, and SBU–YLin, different cloud fraction conditions are shown. WSM6, Goddard, and SBU–YL ~~provide~~ provided a good simulation of the clear sky above the North Sea and Eastern Europe, while the sky above the Alps is as cloudy as in the satellite data. The worst case, Kessler, shows a large cloud cover condition

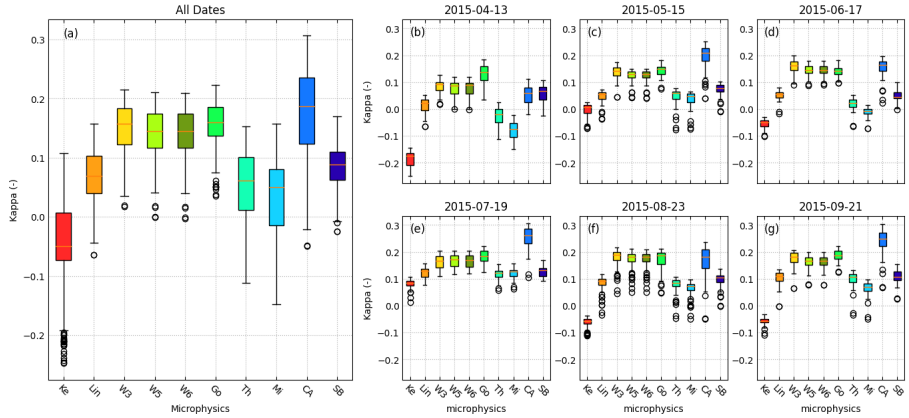

Figure 8. The boxplots of ~~kappa~~ κ in (a) all the simulation dates, (b) case ~~2015-04-13~~April 13, (c) case ~~2015-05-15~~May 15, (d) case ~~2015-06-17~~June 17, (e) case ~~2015-07-19~~July 19, (f) case ~~2015-08-23~~August 23, and (g) case ~~2015-09-21.~~ September 21 of Set 1.

over ~~the~~ Eastern Europe, which contradicts the observation ~~. The high variability of cloud cover causes a less cloudy day in case2015-08-23~~during the dynamic August 23 case.

~~Case 2015-09-21~~ The September 21 case is a cloudy day with low variability in cloud cover. A band of clear sky ~~occur~~ occurs above Austria, Slovakia, southern Poland, and Ukrain (Figure 10). The WSM6 and Goddard microphysics ~~simulate~~ simulated less cloud over the clear sky band but ~~produce~~ produced less cloud cover overall within the model domain. Kessler ~~simulates~~ simulated a cloudy condition over Central Europe.

In Figure 11, the KDE of the cloud cover shows the probabilistic distribution of the average cloud cover for the 36 hours of simulation after hour 12 for ~~case 2015-08-23 and case 2015-09-21. In case 2015-08-23~~the August 23 and September 21 cases. For August 23, CAM5.1 microphysics ~~works~~ worked well for the ~~Kappa~~ κ score but overestimates the cloud cover. The overestimation of cloud cover causes CAM5.1 to perform worse than in ~~cases 2015-05-15, 2015-07-19, and 2015-09-21. Eta,~~ the May 15, July 19, and September 21 cases. WSM3, WSM5, WSM6, and Goddard microphysics ~~shows~~ show similar trends of cloud cover distribution as in the satellite image. In ~~case 2015-09-21~~the September 21 case, the clear-sky condition (< 5% of cloud cover) is captured ~~wellby~~ well by CAM5.1, but the cloud cover distributions of the entire CAM5.1 cluster differ from that in the satellite image.

By comparing the result with the microphysics cluster, PBL physics cluster, and cumulus parameterization cluster, we can ~~see the good performance~~ conclude that there is good performance for cloud simulation by CAM5.1, WSM3, and Goddard microphysics~~, while the~~ . The PBL physics and cumulus ~~paratermeterizations have a secondary impact~~ parameterizations have less impacts on the simulation of cloud cover than the microphysics. To obtain a comprehensive result on cloud cover, we investigate the impact of using different microphysics on the average cloud cover distributions. However, the ~~average~~ temporal- and spatial-averaged cloud covers provide less information and less variability over time. ~~Therefore, the simulated average cloud covers can be performed well and captured by the cloud cover distribution, but the simulation skill can only be~~

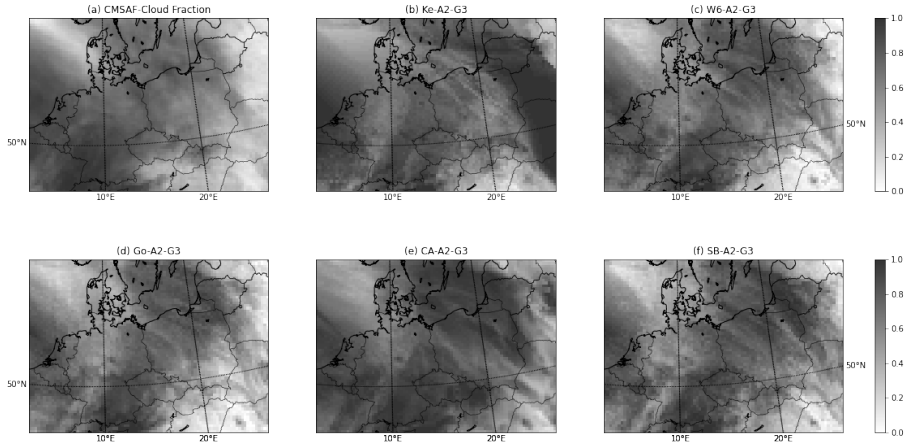

Figure 9. Average cloud cover fraction for 36 hours of (a) satellite data and simulation by different microphysics including (b) Kessler, (c) WSM6, (d) Goddard (e) CAM5.1, and (f) SBU–YLin. All simulations are configured with Grell-3D and ACM2 PBL physics, which ~~performs~~ performed a more skilled prediction than any other combination ~~on 2015-08-23.~~from August 23.

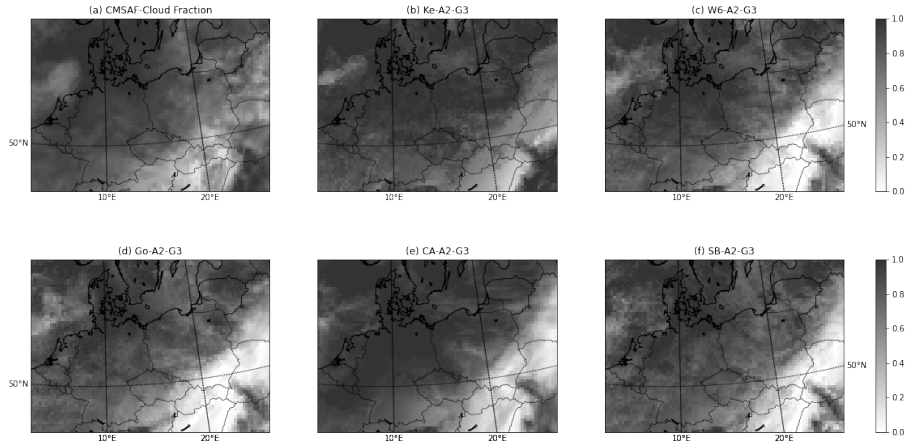

Figure 10. Average cloud cover fraction for 36 hours of (a) satellite data and simulation by different microphysics including (b) Kessler, (c) WSM6, (d) Goddard (e) CAM5.1, and (f) SBU–YLin. All the simulations are configured with Grell-3D and ACM2 PBL physics, which ~~performs~~ performed a more skilled prediction than any other combination ~~on 2015-09-21.~~from September 21.

~~captured pixel-by-pixel with the Kappa score~~ To determine the simulation skill on the spatial patterns, we score the simulation result by calculating the κ score using the pixels in the simulation domain.

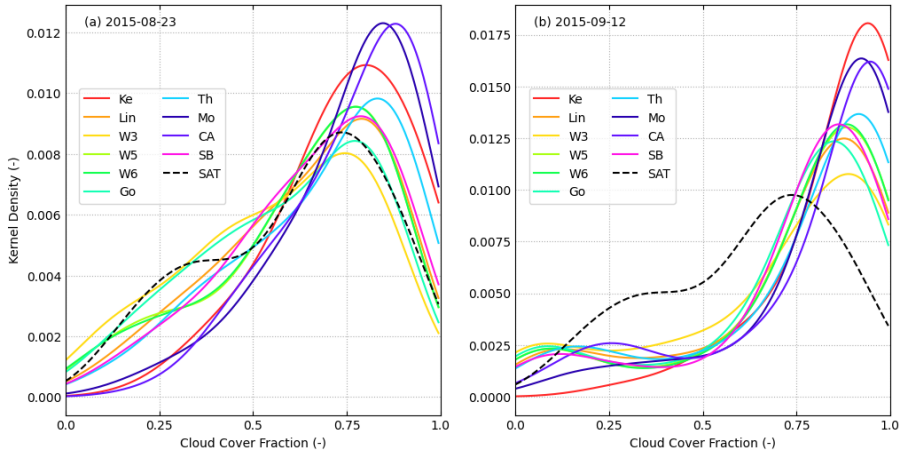

Figure 11. The probabilistic function from the kernel density estimation (KDE) of the average cloud fraction for the last 36 hours simulation for (a) case ~~2015-08-23~~ August 23 and (b) case ~~2015-09-21.~~ September 21. The PBL physics is by ACM2 and the cumulus parameterization is by Grell-3D. The solid color lines depict the KDE from the simulations and the black dash is the KDE from the satellite data

4.3 ~~Sensitivity~~ Set 2 sensitivity analysis~~on the clusters~~ : Clusters of microphysics, PBL, cumulus, radiation schemes, surface layer physics, and land surface parameterization

Additional components to the physics configuration ~~includes~~ include different longwave and shortwave radiation physics and land surface layer physics. However, there are more than ~~1,000,000 combinations from all the~~ a million possible combinations of all physics options. ~~Therefore we~~ We therefore narrow down the choice of the microphysics, PBL physics, and cumulus parameterization from Section 4.2. Accounting for the ~~support of the simulation~~ treatment of the graupel mixing ratio for ESIAS-chem, we predominantly use the microphysics of WSM5, WSM6, and Goddard. CAM5.1 ~~performs~~ performed the best across 5 of the 6 test cases but it is not included because of the ~~high costof simulation resources with the same CPU settings~~higher computational cost. MYNN2, MYNN3, and ACM are selected because of their good performance with the selected microphysics. From the heatmap (Figure 6) both Grell 3D and Tiedtke work well with Goddard and WSM6. We also choose Kain-Fritsch, which is widely used (e.g. Warrach-Sagi et al. (2013) and Knist et al. (2017)), for comparison. The PBL physics ~~by~~ from MYNN2, MYNN3, and ACM2 ~~perform~~ performed well across all the simulations and is chosen for this simulation case.

Figure 12 shows the heat map of the ~~simulation case 2015-09-21~~ September 21 simulation case and the physics configuration of Goddard and ACM2. The Goddard radiation schemes perform skillful predictions of cloud cover. This heat map also indicates good combinations of microphysics, cumulus parameterization, and, radiation schemes as well as combinations of PBL physics, surface physics, and land surface models by row and column, respectively. By row, the Goddard works ~~overall~~ with the Tiedtke and Grell-3D cumulus parameterization over all. By the column, the heat map shows that ACM2 PBL physics can improve the simulation with all the microphysics but with less improvement for the radiation schemes RRTM and Dudhia.

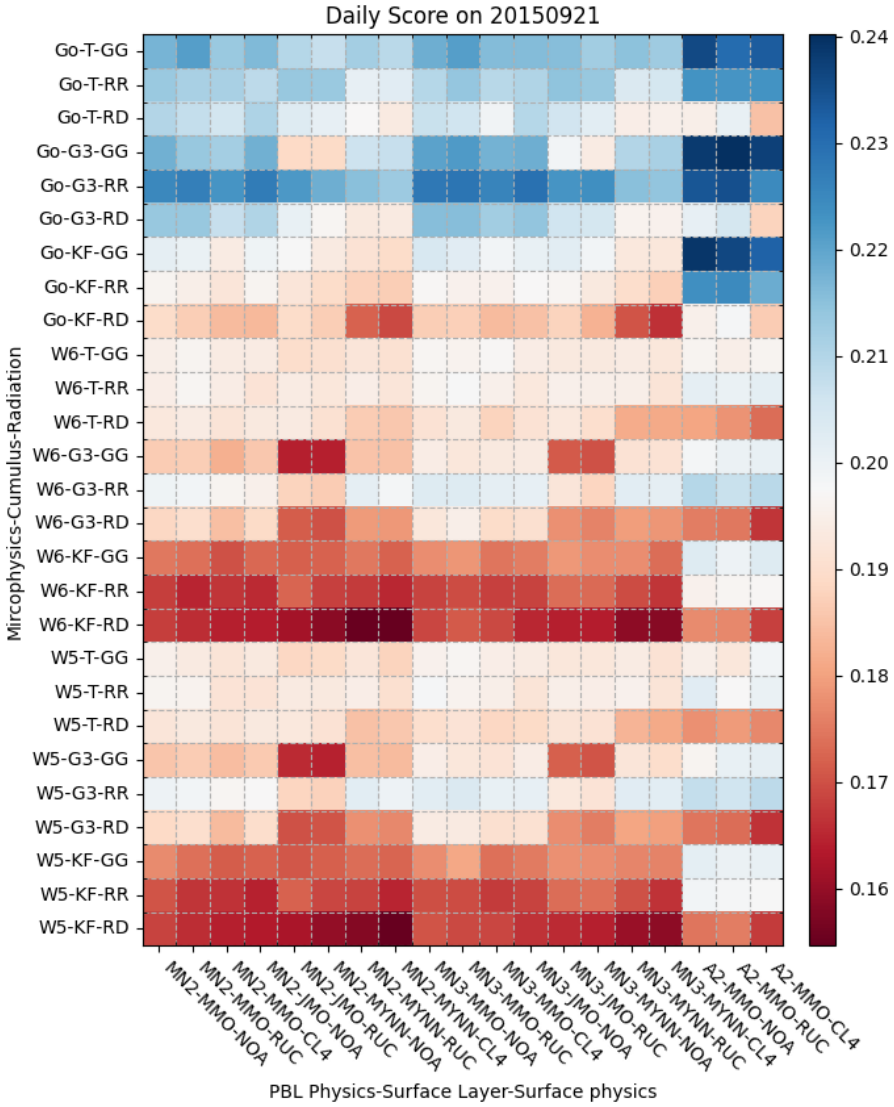

Figure 12. The heat map of the average kappa for the configuration for the clusters of microphysics, cumulus parameterization, and the radiation physics (~~Y axis~~y-axis) with the cluster of PBL-physics, surface physics, and the land surface model (x-axis) on the simulation case ~~2015-09-21.~~ September 21 of Set 2. The data is only use the last 36 hours of simulations for calculating the kappa.

Under the same condition, MYNN3 with Grell-3D and RRTM and Dudhia perform better with different microphysics. From the 513 combinations of physics configurations, the range of the ~~Kappa~~ κ score is between 0.15 and 0.24. The microphysics and the PBL physics are chosen from the results of section 4.2, which is why the improvement is not significant compared to the improvement from changing either the longwave and shortwave radiation scheme ~~,~~ or the surface layer physics ~~, or the~~ or land surface models.

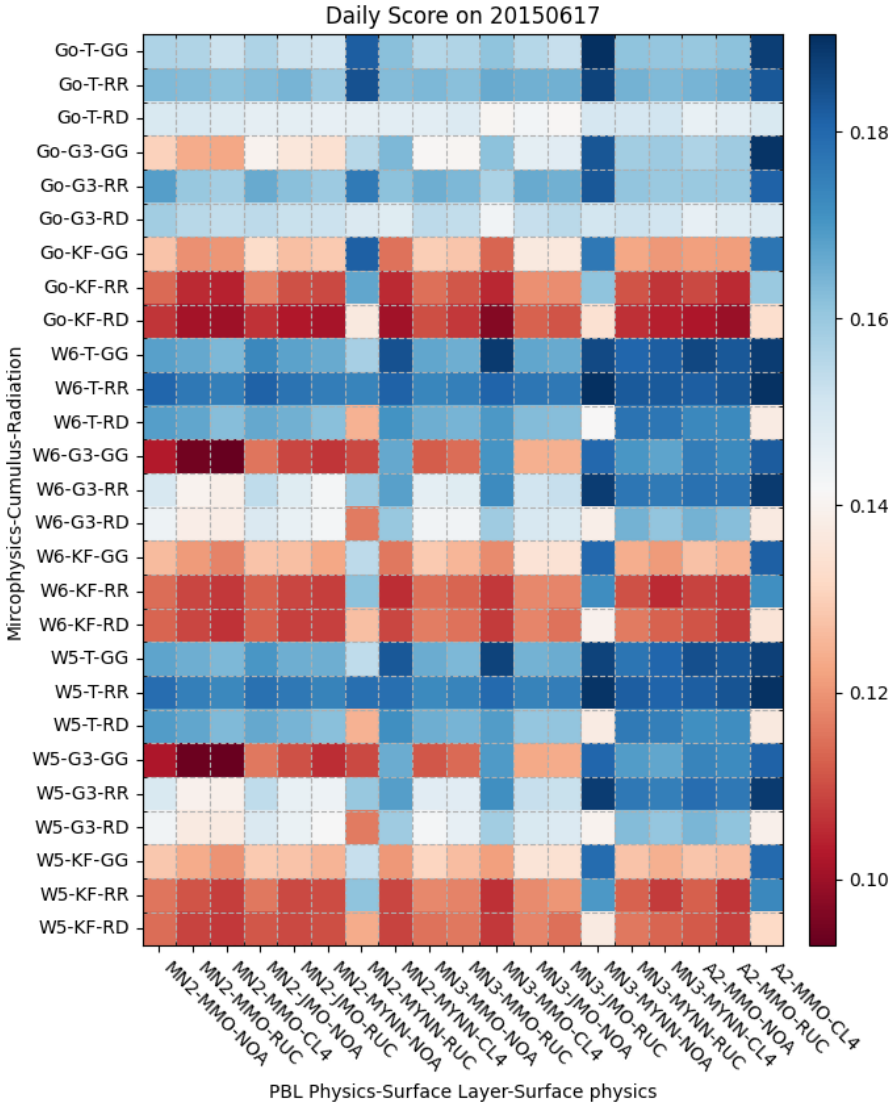

Figure 13. The heat map of the average kappa for the configuration for the clusters of microphysics, cumulus parameterization, and the radiation physics (~~Y axis~~y-axis) with the cluster of PBL-physics, surface physics, and the land surface model (x-axis) on the simulation case ~~2015-06-17.~~ June 17 of Set 2. The data is only use the last 36 hours of simulations for calculating the kappa.

In ~~case 2015-06-17~~the June 17 case, the cloud cover distribution has a very high variability, and therefore the simulation skills increase their variabilities in ~~Kappa~~ κ with different combinations. When all ~~the cluster of microphysics~~ microphysics clusters with Kain-Fritsch ~~perform~~ score less than 0.1, the simulations are better with the ~~combination~~ combinations of MYNN2~~and MYNNand RUC; the combination of MYNN3, MYNN, and Noah; and the combination of ACM2, MMO~~/MYNN/RUC, MYNN3/MYNN/Noah, and ACM2/MMO/CLM4. The pattern of outperforming ~~Kappa~~ κ is also shown in the result for

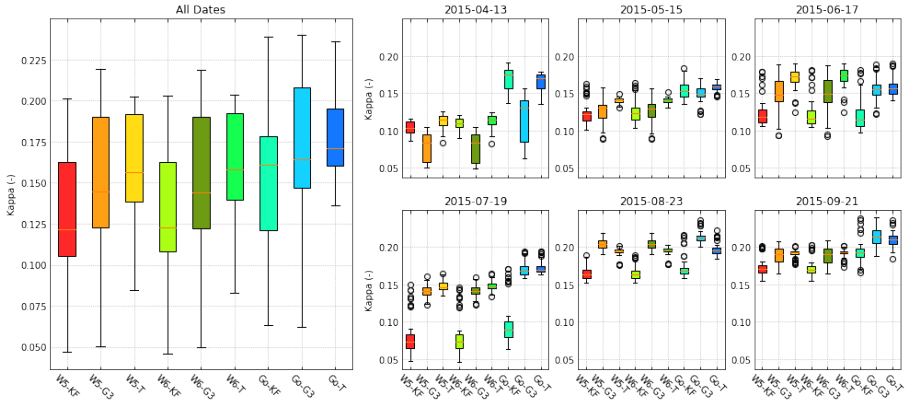

Figure 14. The boxplots of kappa from microphysics and cumulus parameterization clusters in (a) all the simulation dates, (b) case ~~2015-04-13~~April 13, (c) case ~~2015-05-15~~May 15, (d) case ~~2015-06-17~~June 17, (e) case ~~2015-07-19~~July 19, (f) case ~~2015-08-23~~August 23, and (g) case ~~2015-09-21.~~September 21 of Set 2.

~~2015-07-19~~ July 19 (Figure S7), which includes the worst ~~performance of Kappa~~ κ score among the six cases. The results from the other four cases can be found in ~~Figure~~ Figures S5-S8.

The boxplot (Figure 14) shows the overview of the six cases along with the ~~cluster of microphysics and the cumulus parameterizations. The cluster~~ clusters of microphysics and cumulus ~~parameterizations~~ parameterization, which show less variability and are very similar for each case, ~~which indicates~~ indicating that high variability occurs ~~in other clusters of physics. The maximum~~ from other physics clusters. The whiskers of the boxplot (~~3rd quantile + 1.5 × interquartile range) shows that the combination~~ as maximum or minimum of the data, or 3rd quartile $\pm$ 1.5 $\times$ interquartile range) show that combinations with the Grell-3D cumulus parameterization can achieve ~~maximum average Kappa scores. Using~~ the maximum average κ scores. The Goddard microphysics and Tiedke cumulus parameterization can be ~~less variable than the ones using other physics or parameterization, and its median Kappa~~ the least variable, and their median κ indicates that this combination can outperform other microphysics and cumulus parameterization clusters. The boxplots from all the cases show that the cumulus ~~parameterization, which is~~ parameterizations Grell-3D and Tiedtke, which are more advanced than the Kain-Fritsch, can improve the ~~Kappa~~score. The resulting ~~Kappa is not significantly different between using~~ κ does not significantly differ between Grell-3D and Tiedtke.

4.4 ~~Sensitivity~~ Set 3 sensitivity analysis~~on the clusters~~ : Impact of microphysics and PBL ~~for its impact~~ on stochastic simulation

Stochastic weather forecasting ~~depends on a large ensemble size of simulation~~ requires many diverse simulation ensemble members. To study the impact of the physics configuration on the stochastic simulation, we ~~generate~~ generated 31 + 1 ensemble members ~~to simulate 48 hours of stochastic weather forecasting~~in 48-hour ensemble runs. The total cloud fractions again after 12 simulation hours are used to ~~analyse the effect of different physics configurations on the simulation result and its impact on~~

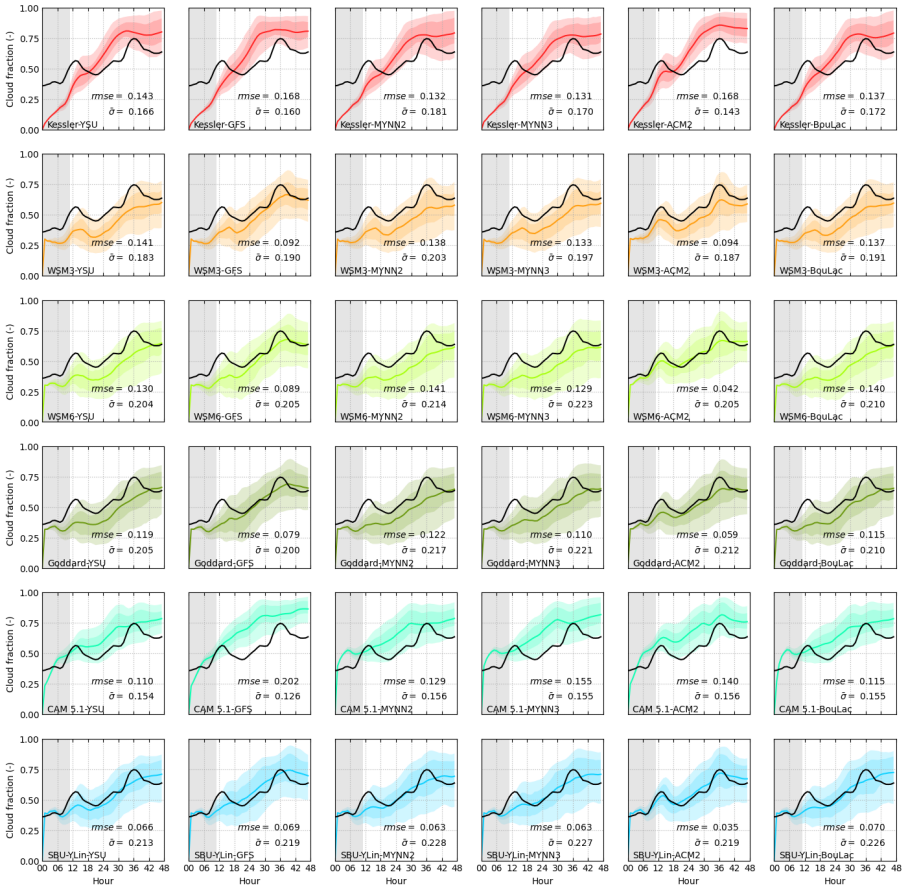

Figure 15. The mean cloud cover fraction (-) the observation from the satellite (black line) and by simulation (color line) from the combination of microphysics (by different color) and PBL physics (in different row) in case ~~2015-08-23.~~ August 23. The color block represents the range of percentiles, the darker block is limited between 25% and 75%, and the lighter block is lmited between 5% and 95%. The grey block indicates the spin-up time for ESIAS-met, which is not included in the root mean square error *rmse*, standard deviation ($\bar{\sigma}$), and the mean simulated cloud cover fraction ($\bar{x}$).

~~the probability. The simulation is conducted for the six cases with the same domain setting and input data and the~~ analyze the differences from the model configuration and their impact on probabilities. The stochastic experiments are simulated for the same ~~cloud data are also used~~cases and domain and with the same input data as before.

Figure 15 shows the probabilistic cloud cover fraction within the 25th to 75th percentiles and 5th to 95th percentiles of the simulations. The ~~developement~~ development of the mean cloud cover fraction ~~are~~ is compared to the mean cloud cover fraction of the satellite data. The *rmse* and standard deviation are used to ~~the~~ show the comprehensive result ~~from the temporal development of cloud cover within the last~~ of the temporal cloud cover development within the final 36 hours of ~~simulation~~the simulations.

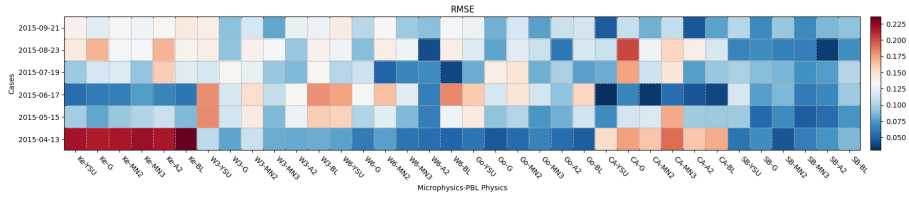

Figure 16. ~~The~~ Heat map of the RMSE between the ensemble mean total cloud ~~cover~~ fraction ~~(-) the~~ and observation~~from~~, averaged over the ~~satellite (black line) and by simulation (color line) from~~ last 36 hours of the ~~combination~~ simulations, shown for each cluster of microphysics ~~(by different color)~~ and PBL physics (~~in different row~~columns) ~~in case 2015-09-21. The color block represents the range of percentiles, the darker block is limited between 25% and 75%, and the lighter block is limited between 5% and 95%. The grey block indicates the spin-up time~~ for ~~ESIAS-met which is not included in the root mean square error *rmse*, standard deviation~~ all test cases ($\bar{\sigma}$rows). In the colorscale, red represents higher RMSE and ~~the mean simulated cloud cover fraction ($\bar{x}$)~~poorer performance.

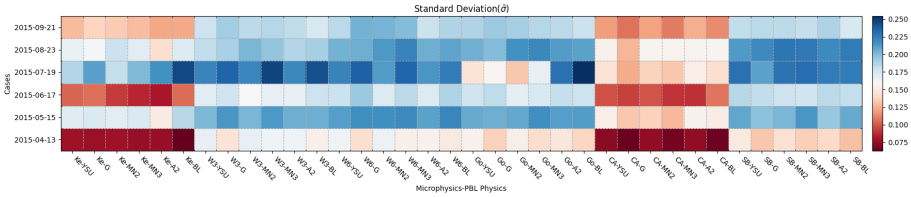

Figure 17. Heat map of the ensemble standard deviation ($\bar{\sigma}$) of the mean total cloud fraction, averaged over the last 36 hours of the simulations, shown for each cluster of microphysics and PBL physics (columns) for all test cases (rows). In this heat map, blue indicates larger $\bar{\sigma}$ and greater ensemble spread.

The Kessler microphysics ~~show an overestimation of~~ appeared to overestimate the cloud cover fraction and have the largest *rmse* of all microphysics. Moreover, the two peaks ~~of~~ in the cloud cover fraction ~~are not captured at~~ around the 12th hour and ~~the~~ 36th hour of simulation ~~. In addition to Kessler microphysics, the microphysics cluster can capture these two peaks with the~~ are not clearly captured by the Kessler ensemble, though this is better than in most cases in combination with ACM2~~physics. The smallest *rmse* is performed by~~ . The SBU–Ylin microphysics and ACM2 PBL physics ~~and the second smallest by~~ achieved the smallest *rmse* and the WSM6 microphysics and ACM2 PBL physics were second best. The biggest standard deviation ~~is produced by~~ and therefore ensemble spread was produced by the SBU–Ylin and MYNN2 PBL physics~~, and this physics configuration therefore produces the largest spread of possible results~~. In the ~~2015-09-21~~ September 21 case (Figure ~~??~~S13, the WSM6 microphysics ~~shows~~ showed the greatest spread ~~with the~~ and largest standard deviation. In both cases, the CAM5.1 microphysics combined with the GFS PBL physics ~~produces the smallest~~ to produce the narrowest probability distribution of the cloud cover fraction. The color blocks show the simulation skill in capturing the cloud cover fraction within a certain percentile. The MYNN2, MYNN3, and ACM2 PBL physics not only produce a larger probabilistic distribution than the other PBL physics, but also perform better ~~results~~ with the WSM6, Goddard, and SBU–YLin microphysics. The results from the other four test cases can be found in ~~Figure~~ Figures S9-S12.

5 ~~Discussion and conclusion~~

Figure 16 summarizes the rmse performance of the ensemble mean against the observed domain total cloud fraction. Figure 17 illustrates the time-averaged ensemble spread with the standard deviation (std) of the domain total cloud fraction. From the rmse, SBU–YLin had the best mean total cloud fraction over all six cases, with a more variable cloud fraction according to std. The WSM3, WSM6, Goddard, and SBU–YLin with MYNN3 and ACM2 produced more accurate average cloud fractions than the other combinations. Overall, the WSM series and SBU–YLin better represented the uncertainty than the other microphysics, while MYNN3 and ACM2 improved the simulation accuracy.

4.1 ~~Impact of physics configuration on the simulated cloud cover~~Set 4 sensitivity analysis: Long-term validation

Sets 1-3 address over a thousand model combinations. Ideally, all combinations could be tested over a full year to capture diverse conditions and seasonality and to benchmark their operational performance. The aggregate computational expense of all combinations is unfortunately too great for this. For the four most promising models listed in Table 4, however, ESIAS could be economically tested over a full year and these four cases compared to roughly 6 months of available satellite data in 2018. A summary of these results is shown in Figure 18.

In this test of the more generalized model performance, the models scored similarly, with the WSM6 microphysics having higher matching rates and lower standard deviations than Goddard with the same PBL, and likewise, MYNN3 performing better than MYNN2 for the same microphysics. In the plot, the maximum matching rates are thus generally reached by points with MYNN3, especially with WSM6, whereas the lowest are most commonly from Goddard-MYNN2.

The hourly matching rates can fluctuate up to about 10% within a given day, but also at timescales of weather patterns over 2-4 weeks, generally around their means of $\sim 78\%$. There is no clear seasonal trend in the matching rate over these six months. The models perform relatively better or worse on the same days, depending on the weather condition, as they all share the same ECMWF input data and the WRF outputs remain similar at the synoptic scale. The model preferences here seem consistent with the relative performances in Sets 1 & 2 for the 2015 test days, such that these four configurations seem suitable for the general simulation of cloud cover with WRF on the European domain.

5 Discussion

5.1 Impact of physics configuration on the simulated cloud cover

The ~~Kappa heatmap from the~~ cloud cover masks ~~shows that microphysics is the key aspect affecting~~ κ heatmap shows that the choice of microphysics is most consequential to the simulation of cloud cover in the European domain. The ~~Kappa~~ values show a consistently good result ~~by using the~~ from WSM3, WSM5, WSM6, Goddard, and CAM5.1 microphysics based on the six ~~simulation cases and 672 combination of physics configuration (~~test cases and 560 physics combinations in Set 1 ~~)~~ and 513 ~~combination of physics configuration (Set 2).~~ combinations in Set 2. The good performance ~~in predicting~~ of WSM6 for cloud cover fraction ~~using WSM6 had~~ has been previously explained by Pieri et al. (2015) and Jankov et al. (2011).

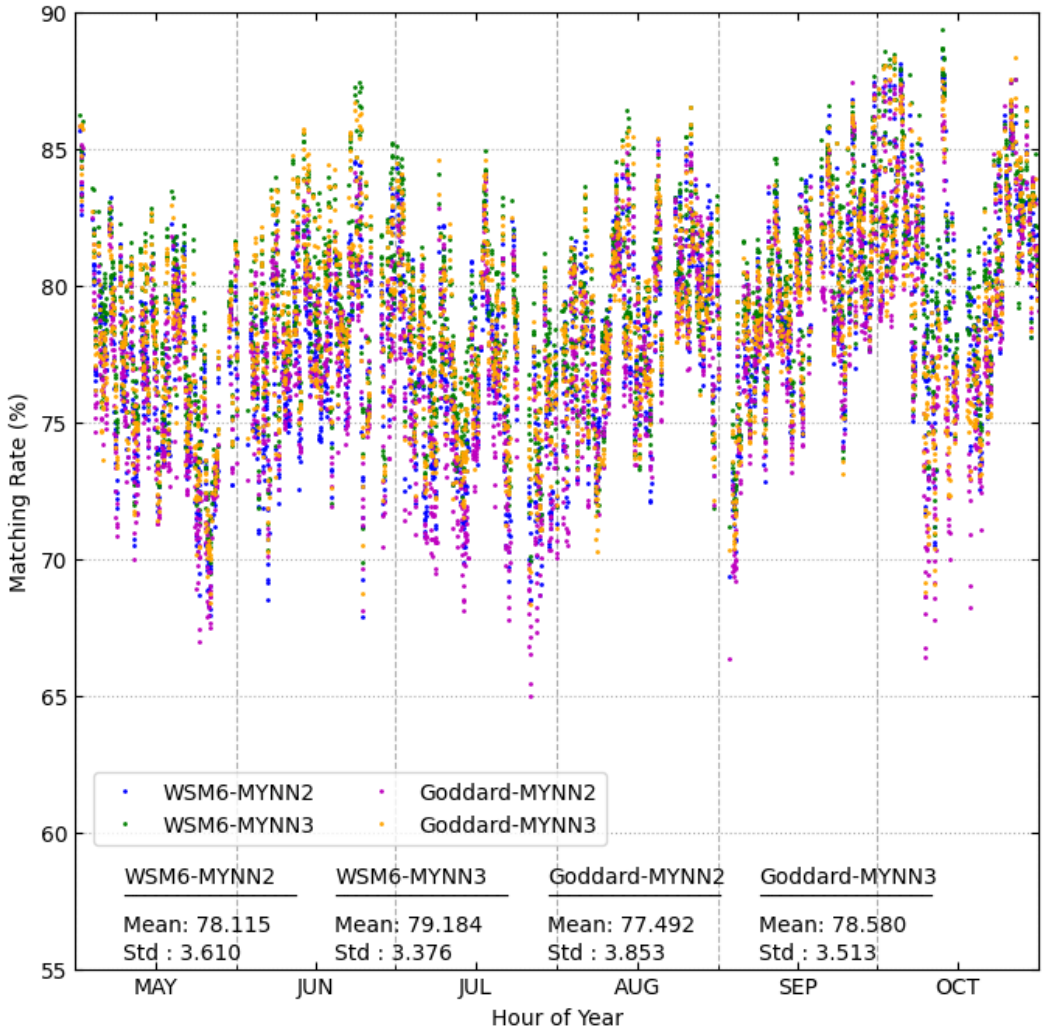

Figure 18. Six-month time series of the hourly matching rate the four Set 4 combinations. The means and standard deviations of the matching rates are shown beneath the legend.

The employment of ACM2, MYNN2, ~~and~~ or MYNN3 PBL physics can lead to good results in the cloud cover mask. ~~For the six cases the Morrison~~ According to the six test cases, the Milbrandt 2-mom and Kessler microphysics schemes should be avoided, ~~and~~ as should the QNSE and GFS PBL physics~~should also be avoided~~. The research results of Borge et al. (2008) point to the same choice of WSM6 microphysics, but our results ~~instead suggest using~~ differs in suggesting ACM2, MYNN2 and MYNN3 PBL physics, while Borge et al. (2008), Gbode et al. (2019) and Stergiou et al. (2017) ~~report the~~ rather recommend YSU or MYJ PBL physics~~to be the better physics configuration. However, Borge et al. (2008) focus~~ . Borge et al. (2008) however focuses on the prediction of wind properties, temperature, and humidity over Spain for 168 hours of simulation~~;~~

~~Gbode et al. (2019) focus~~ , Gbode et al. (2019) focuses on the prediction of precipitation over West Africa during the monsoon period, and Stergiou et al. (2017) focus on the prediction of temperature and precipitation over Europe during two different months. This study focuses on cloud cover and the solar power application.

~~The boxplots by~~ Boxplots of the different physics combinations ~~illustrate~~ illustrated that the employed land surface model ~~or radiation schemes are not as consequential as the change of configuration for~~ and radiation scheme was less consequential than the microphysics and PBL physics. However, the results from the Set 2 sensitivity analysis showed significant differences for different cumulus parameterizations, ~~although~~ though not as large as from the change of the microphysics and PBL physics, and the ~~Kappa values are~~ κ values were generally less for the Kain-Fritsch ~~than from~~ model than for Grell-3D and Tiedtke.

5.2 Impact of physics configuration on ~~the result of~~ stochastic simulation

As ESIAS-met is ~~the~~ an ensemble version of WRF, ~~and therefore it is very~~ it was therefore important to understand the impact of different physics ~~configuration on the~~ on stochastic results. We ~~identify~~ identified the most sensitive physics clusters and ~~use the physics configuration (~~used Set 3 ~~)~~ to generate 1,~~536 ensemble members . However, we only simulate 31 + 1 ensemble members due to the limitation of computation~~ 152 total members from 36 physics × 32–members with the SKEBS scheme. We were limited to ensemble groups of this size due to limited computational resources (i.e. a single case uses 18,432 cores for a total of originally 1,536 ensemble members, from which 192 were excluded due to unsuitable microphysics).

The ~~results of the stochastic simulations~~ stochastic results show that the combination of physics highly affects the probability distribution. Kessler ~~overestimates~~ overestimated the cloud cover fraction~~. Similar to~~ , while as in the results from the ~~physics configuration sensitivity analysis~~preceding sets, the cloud cover prediction ~~is simulated well by~~ was simulated well with the microphysics CAM5.1, Goddard, and WSM6. Moreover, the SKEBS stochastic scheme ~~produces~~ produced broader probabilistic distributions ~~by employing~~ with WSM6 and Goddard and ~~therefore~~ the average cloud cover fraction ~~can be captured~~ could therefore be captured best by these microphysics. The CAM5.1 microphysics ~~produces~~ produced the most accurate results when compared pixel by pixel, but the probabilistic distribution ~~is smallest compared to the other microphysics~~ was the smallest of the microphysics options.

~~The PBL physics not only change~~ Likewise, the PBL physics affected not only the development of the cloud cover fraction but also ~~affect~~ the probabilistic distribution~~of the cloud cover fraction~~. The GFS and MYNN2 ~~scheme produce~~ schemes produced less dynamic cloud cover and thus ~~produce~~ higher *rmse* values. ~~The~~ ACM2 ~~produces~~ produced a more dynamic development of cloud cover, but its probabilistic distribution ~~is~~ was slightly less than that of MYNN2 and MYNN3.

~~Stochastic~~ The stochastic analysis shows a contradiction between ~~the deterministic simulation and the~~ deterministic accuracy and probabilistic simulation. The most ~~suitable configuration for an unbiased~~ accurate configuration for a deterministic forecast may differ from that ~~of the best multi-member ensemble forecast~~ for the ensemble with the most accurate mean or that best captures the uncertainty and diversity of ~~the~~ possible outcomes. ~~For the most accurate simulation, the~~ The CAM5.1 microphysics and ~~the~~ ACM2 PBL physics ~~performed best~~lead to the most accurate deterministic forecast as compared to the satellite observation, while SBU–YLin with MYNN2 ~~shows better results~~ was the most accurate ensembles in terms of ~~mean cloud cover and the observed satellite data. Of all the~~ its mean. Through all six cases, the CAM5.1 microphysics ~~produces the least~~

~~probabilistic~~ produced the narrowest distribution, while the Goddard and WSM6 microphysics ~~can~~ could generally produce broader probabilistic distributions.

~~Jankov et al. (2019)~~ Jankov et al. (2017) and Li et al. (2019) both ~~report an insufficiency of~~ reported insufficiencient ensemble spread with stochastic schemes (e.g. SKEBS or SPPT) and ~~suggest using a multiphysics simulation to obtain~~ that mixing multiple physics in simulations can achieve a greater spread. In our simulations, ~~the multi-physics can enhence the spread of ensemble simulation, though the ensemble spreadby multi-physics simulation is based on the uncertainty in the model physics. A probabilistic simulation is a solution that predicts weather forecasting from the distribution of large ensemble member sets, but we should also consider the accuracy~~ combining the ensembles into one multi–physics ensemble would enhance the spread, but this would be somewhat artificially due to the different biases of the model physics. The ~~multi-physics simulation serves as an ensemble simulation based on random estimation. For instance,~~ accuracies of the different model physics ~~simulate different cloud patterns, and the sum of two cloud patters then eliminate cloud boundaries and are not in agreement on the resulting front. Moreover, Jankov et al. (2019) performed a stochastic simulation with only four ensemble members; the insufficient number of ensemble members may be the root cause of the small ensemble spread reported in their results~~must then always been considered for multi-physics ensembles. We also note that the small ensemble spread reported in Jankov et al. (2017) may be due to the small number of ensembles, four for each physics configuration, yielding eight members in total.

5.3 Choice of physics configurations

The simulation results do not indicate a single best option for the physics configuration. Many studies ~~focusing~~ exist that focus on very different ~~topics, including different resolutions~~ aspects of sensitivity analysis, including spatial resolution (Warrach-Sagi et al., 2013; Pieri et al., 2015; Knist et al., 2017, 2018), inputs (Pieri et al., 2015), microphysics (Jankov et al., 2011; Rögnvaldsson et al., 2011), PBL physics (García-Díez et al., 2013), cumulus parameterizations (Gbode et al., 2019) , land surface models (Jin et al., 2010) , and the combination of different physics (Borge et al., 2008; Santos-Alamillos et al., 2013; Awan et al., 2011; Jankov et al., 2007; Pieri et al., 2015; Stergiou et al., 2017; Otkin and Greenwald, 2008; Li et al., 2019; Varga, 2020). ~~However, these~~ These studies focus on different target variables and meteorological states with different weather forcing input, observation data, ~~study areas~~domains, and time scales, ~~which~~ and therefore produce very different results for the choice of physics or parameterization. Our simulation results ~~do not give a clear indication on the meteorological aspect~~ cannot give a clearer indication of the meteorological aspects across temporal and spatial scales~~. Therefore we offer a recommendation on the choice of~~ , but can suggest some best physics configurations for studying cloud simulation or solar power over the European domain~~and for weather forecasting purposes. However, further investigations must~~ . Further investigations must still be carried out for more comprehensive insights into on the spatial scales, ~~more combinations of~~ meteorological variables, further physics configurations, and different input data (e.g. ECMWF ERA5).

Nevertheless, ~~the simulation applying~~ regarding day-ahead simulations with 20 km horizontal resolution~~with the day-ahead weather forecasting, and the employment of~~ , WSM6, Goddard, and CAM5.1 microphysics ~~improve~~ performed best here for deterministic weather forecasting ~~on cloud cover, while the employment of~~ of cloud cover. In the probabilistic application, WSM6, Goddard, and SBU–YLin microphysics ~~benefit the stochastic weather forecasting on cloud cover with a greater probabilistic~~

~~distribution~~yielded the greatest variability, while Kessler and CAM5.1 conversely generated the narrowest distributions. The PBL physics ~~are better~~ were best simulated by ACM2 and MYNN3. ~~However, cumulus parameterizations~~The choice of cumulus parameterization, surface layer physics, and ~~the land surface models do not increase the accuracy significantly . A more comprehensive study should include some promising combinations of physics configurations, and include the short- or long-term effects of applying different physics at different spatial scales, such as continental or global scales, and include a 1 km resolution to study the dynamics and local conditions at convection-resolving scale.~~

~~The performance of probabilistic simulations relies on their probabilistic distribution in addition to their~~ land surface model did not significantly increase the accuracy. The ~~SBU−YLin, Goddard, and WSM6 schemes generate broader probabilistic distributions, while Kessler and CAM5.1 generate the narrowest. The PBL scheme also has~~ PBL scheme did have a significant effect on the probabilistic distributions~~;~~. MYNN2 and MYNN3 ~~generate~~ generated wider distributions while GFS ~~generates~~ generated smaller ones.

The best combination for ~~probabilistic simulations~~ ensemble-based probabilistic simulation ideally has accurate ensemble means and realistically broad distributions. The Goddard, WSM6, or SBU–YLin microphysics with MYNN3 are potential choices. When the mean of the simulations is close to the mean of the observation, the reality can be better captured. However, as mentioned by Sillmann et al. (2017), the technique of scoring ensemble simulations remains a challenge ~~in better estimating the results from the view of~~ for better probabilistic analysis. A more comprehensive study should also include some promising physics combinations while including the short- or long-term effects of applying different physics at different spatial scales, such as continental or global scales, as well as including a 1 km resolution to study the dynamics and local conditions at the convection-resolving scale.

5.4 Future work

~~We performed simulations~~ This study was performed without a nested domain for a higher resolution simulation, which might be useful for investigating the effect of the resolution on multi-physics for convection-resolving simulations. Exascale high-performance computing might enable ~~make~~ such studies for scientific research and provide an opportunity for investigating the scalability of ultra-large ensemble simulation systems (Neumann et al., 2019; Bauer et al., 2021). To this end, the ESIAS system presented here has been designed to perform data assimilation with the advantage of ~~the~~ its elastic ensemble simulation frameworks. Further development will focus on data assimilation, such as the use of the particle filter with the particle removal function (van Leeuwen and Jan, 2009).

6 Conclusions

This study ~~introduces~~ introduced an ensemble simulation system for conducting ultra-large ensemble simulations in Europe and with multiphysics ~~simulations and~~ and ensemble-based probabilistic simulations. We ~~use~~ used the meteorological part of the system to ~~perform simulations and generate large ensembles to perform sensitivity analyses on the effect~~ generate diverse simulations and perform and iterative sensitivity analyses of various physics configurations ~~on the~~ for cloud cover fraction ~~.~~

~~The simulation is conducted within Europe without any nested domain for a day-ahead forecasting (48 hours) for six days. The sensitivity analysis is based on 672 and~~ using κ coefficients to score the match rate of cloud cover masks. This began with experiments on 6 days of 560 initial physics combinations, followed by 513 ~~combinations of physics configurations, investigating combinationsof three and six different physics elements, respectively. Additionally, we perform a sensitivity analysis to determine the effect of physics configurations on cloud fraction and use the calculation of Kappa to identify the score of the match rate by cloud cover masks~~additional tests of secondary model choices, and finally tests of the stochastic performance of 42 selected combinations. Lastly, a half year of data could be simulated to test the long-term performance of the four favored model combinations.

The sensitivity analysis of the combination of three physics configurations – microphysics and the planetary boundary layer (PBL) physics and the cumulus parameterization – ~~shows that the microphysics has~~ showed the microphysics to have the greatest influence on ~~the~~ cloud cover. The Goddard, WSM3, and CAM5.1 microphysics consistently ~~perform~~ performed better than the other microphysics, but the amount of computation time required for CAM5.1 is relative high. The Goddard and WSM3 scheme ~~increase performance when the cloud cover is more dynamics~~did better for more dynamic cloud situations. The PBL physics also ~~has~~ have a significant effect on the results and ~~shows~~ show better agreement with ~~YS~~ MYNN2/3 and ACM2, but less agreement with GFS and QNSE PBL physics. The long-term simulation using WSM6 and Goddard with MYNN2/3 in 2018 showed that the agreement between simulated and observed cloud mask reaches at least 65% and at largest 89% without trends in different seasons.

The sensitivity analysis on the combination of six physics configurations – including microphysics and the PBL physics, the cumulus parameterization, longwave and shortwave radiation schemes, surface layer physics, and land surface models – shows again that the microphysics ~~affects~~ affect the cloud cover the most, and that the ACM2 PBL physics significantly increases the ~~accuracy of predicting cloud cover . However, the~~ cloud cover prediction accuracy. The physics configurations of surface layer physics and land surface models ~~are not as significant as~~ were found to be less significant than other physics.

The sensitivity analysis ~~on the combination of two physics configurations and its effect on stochastic simulation shows a significant effect on the probabilistic distributions on the cloud cover fractions.~~ for stochastic simulation showed significant differences. The WSM6 and SBU–YLin microphysics with MYNN2 and MYNN3 ~~capture~~ captured the cloud fraction better within ~~the greater range of their~~ their broader probabilistic distributions than ~~have the~~ did other models, although the WSM6 and ~~SBU-YLin~~ SBU–YLin with ACM2 better ~~captures~~ captured the dynamics of ~~cloud fractions when the cloud clover has more variability during the simulation~~ the cloud fraction in situations with more variability of the cloud cover in time.

The simulation results indicate a pathway for improving model physics and demonstrate the potential of ultra-large ensemble simulations and high-performance computers approaching exascale. The multi-physics simulation however produces a larger ensemble spread compared to the stochastic schemes, although the result from the sum of the multi-physics may not be realistic. The employment of ~~ultra-large ensemble simulations with~~ suitable physics configurations can improve both the accuracy and the probabilistic ~~distributions from simulation results~~quality of both deterministic and large ensemble weather predictions.

Code availability.

The codes of ESIAS-met and also the pre- and post-processed codes are available for public domain via https://zenodo.org/record/6637315#.YqbhgBxBzeK (DOI:10.5281/zenodo.6637315). The modelling and analysis tools can be found in the code repository: https://github.com/hydrogencl/WRF_TOOLS and https://github.com/hydrogencl/SciTool_Py.

Appendix A

~~Abbreviation of the physics employed in this study for microphysics, cumulus parameterization, and planetary boundary layer (PBL) physics. Microphysics Cumulus Parameterization PBL physics Full name Abbr. Full name Abbr. Full name Abbr. Kessler Ke Kain-Fritsch KF YSU YSU Lin (Purdue) Lin Betts-Miller-Janjic BM MYJ MY WSM3 W3 Grell-Freitas GF GFS G WSM5 W5 Simplied Arakawa-Schubert OS QNSE Q Eta (Ferrier) Eta Grell-3 G3 MYNN2 MN2 WSM6 W6 Tiedtke T MYNN3 MN3 Goddard Go New SAS NS ACM2 A2 Thompson Th BouLac BL Milbrandt 2-mom Mi Morrison 2-mom Mo CAM 5.1 Ca SBU–YLin SB~~

~~Abbreviation of the physics employed in this study for the combination of shortwave and longwave radiation scheme, surface layer physics, and land surface model. Radiation scheme Surface layer physics Land surface model (shortwave & longwave) Full name Abbr. Full name Abbr. Full name Abbr. RRTM & Dudhia RD Revised MM5 Monin-Obukhov MMO unified Noah YSU RRTMG & RRTMG RR Monin-Obukhov (Janjic Eta) EMO RUC MY New Goddard & New Goddard GG MYNN MYNN CLM4 G~~

Author contributions.

YSL and GG designed the experiments and wrote the manuscript, and YSL performed the simulation. HE is the founder of the idea of ESIAS-met and helps to prepare of the manuscript.

Competing interests.

The authors declare no conflicts of interest with respect to the results of this manuscript.

Acknowledgements. This work was fully supported by the Council of the European Union (EU) under the Horizon 2020 Project "Energy Oriented Center of Excellence: toward exascale for energy" - EoCoE II, Project ID 824158. The authors also gratefully acknowledge the Earth System Modelling Project (ESM) for funding this work by providing computing time on the ESM partition of the supercomputer JUWELS at the Jülich Supercomputing Centre (JSC).

References

User's Guide for the Advanced Research WRF (ARW) Modeling System Version 3.7, https://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3/user_guide_V3.7/users_guide_chap5.htm, 2015.

Adeh, E. H., Good, S. P., Calaf, M., and Higgins, C. W.: Solar PV Power Potential is Greatest Over Croplands, Scientific Reports, 9, 11 442, https://doi.org/10.1038/s41598-019-47803-3, 2019.

Awan, N. K., Truhetz, H., and Gobiet, A.: Parameterization-Induced Error Characteristics of MM5 and WRF Operated in Climate Mode over the Alpine Region: An Ensemble-Based Analysis, J. Climate, 24, 3107–3123, https://doi.org/10.1175/2011jcli3674.1, https://doi.org/10.1175/2011JCLI3674.1, 2011.

Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, Nature, 525, 47–55, https://doi.org/10.1038/nature14956, https://doi.org/10.1038/nature14956, 2015.

Bauer, P., Dueben, P. D., Hoefler, T., Quintino, T., Schulthess, T. C., and Wedi, N. P.: The digital revolution of Earth-system science, Nature Computational Science, 1, 104–113, https://doi.org/10.1038/s43588-021-00023-0, https://doi.org/10.1038/s43588-021-00023-0, 2021.

Bellprat, O., Guemas, V., Doblas-Reyes, F., and Donat, M. G.: Towards reliable extreme weather and climate event attribution, Nature Communications, 10, 1732, https://doi.org/10.1038/s41467-019-09729-2, 2019.

Bentley, J. L., Stanat, D. F., and Williams, E.: The complexity of finding fixed-radius near neighbors, Information Processing Letters, 6, 209–212, https://doi.org/https://doi.org/10.1016/0020-0190(77)90070-9, https://www.sciencedirect.com/science/article/pii/0020019077900709, 1977.

Berndt, J.: On the predictability of exceptional error events in wind power forecasting–an ultra large ensemble approach–, 2018.

Berner, J., Shutts, G. J., Leutbecher, M., and Palmer, T. N.: A Spectral Stochastic Kinetic Energy Backscatter Scheme and Its Impact on Flow-Dependent Predictability in the ECMWF Ensemble Prediction System, J. Atmos. Sci., 66, 603–626, https://doi.org/10.1175/2008jas2677.1, https://doi.org/10.1175/2008JAS2677.1, 2009.

Berner, J., Ha, S.-Y., Hacker, J. P., Fournier, A., and Snyder, C.: Model Uncertainty in a Mesoscale Ensemble Prediction System: Stochastic versus Multiphysics Representations, Mon. Wea. Rev., 139, 1972–1995, https://doi.org/10.1175/2010mwr3595.1, https://doi.org/10.1175/2010MWR3595.1, 2011.

Borge, R., Alexandrov, V., José del Vas, J., Lumbreras, J., and Rodríguez, E.: A comprehensive sensitivity analysis of the WRF model for air quality applications over the Iberian Peninsula, Atmospheric Environment, 42, 8560–8574, http://www.sciencedirect.com/science/article/pii/S1352231008007954, 2008.

Buizza, R., Milleer, M., and Palmer, T. N.: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system, Q.J.R. Meteorol. Soc., 125, 2887–2908, https://doi.org/10.1002/qj.49712556006, https://doi.org/10.1002/qj.49712556006, 1999.

Dai, A., Meehl, G. A., Washington, W. M., Wigley, T. M. L., and Arblaster, J. M.: Ensemble Simulation of Twenty-First Century Climate Changes: Business-as-Usual versus CO2 Stabilization, Bull. Amer. Meteor. Soc., 82, 2377–2388, https://doi.org/10.1175/1520-0477(2001)082<2377:esotfc>2.3.co;2, https://doi.org/10.1175/1520-0477(2001)082<2377:ESOTFC>2.3.CO;2, 2001.

Diaz, J., Cheng, C.-H., and Nnadi, F.: Predicting Downward Longwave Radiation for Various Land Use in All-Sky Condition: Northeast Florida, Advances in Meteorology, 2014, 525 148, https://doi.org/10.1155/2014/525148, https://doi.org/10.1155/2014/525148, 2014.

Dixon, B. and Candade, N.: Multispectral landuse classification using neural networks and support vector machines: one or the other, or both?, null, 29, 1185–1206, https://doi.org/10.1080/01431160701294661, https://doi.org/10.1080/01431160701294661, 2008.

Dupuy, F., Lu, Y.-S., Good, G., and Zamo, M.: Calibration of solar radiation ensemble forecasts using convolutional neural network, in: EGU General Assembly Conference Abstracts, pp. EGU21–7359, 2021.

Ehrendorfer, M.: Predicting the uncertainty of numerical weather forecasts: A review, METEOROLOGISCHE ZEITSCHRIFT-BERLIN-, 6, 147–183, 1997.

Fleiss, J. L. and Cohen, J.: The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability, Educational and psychological measurement, 33, 613–619, 1973.

Franke, P.: Quantitative estimation of unexpected emissions in the atmosphere by stochastic inversion techniques, 2018.

Franke, P., Lange, A. C., and Elbern, H.: Particle-filter-based volcanic ash emission inversion applied to a hypothetical sub-Plinian Eyjafjallajökull eruption using the Ensemble for Stochastic Integration of Atmospheric Simulations (ESIAS-chem) version 1.0, Geoscientific Model Development, 15, 1037–1060, https://doi.org/10.5194/gmd-15-1037-2022, https://gmd.copernicus.org/articles/15/1037/2022/, 2022.

García-Díez, M., Fernández, J., Fita, L., and Yagüe, C.: Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe, Q.J.R. Meteorol. Soc., 139, 501–514, https://doi.org/10.1002/qj.1976, https://doi.org/10.1002/qj.1976, 2013.

Gbode, I. E., Dudhia, J., Ogunjobi, K. O., and Ajayi, V. O.: Sensitivity of different physics schemes in the WRF model during a West African monsoon regime, Theoretical and Applied Climatology, 136, 733–751, https://doi.org/10.1007/s00704-018-2538-x, 2019.

Gneiting, T. and Raftery, A. E.: Weather Forecasting with Ensemble Methods, Science, 310, 248, http://science.sciencemag.org/content/310/5746/248.abstract, 2005.

Hamill, T. M., Whitaker, J. S., Fiorino, M., and Benjamin, S. G.: Global Ensemble Predictions of 2009's Tropical Cyclones Initialized with an Ensemble Kalman Filter, Monthly Weather Review, 139, 668–688, https://journals.ametsoc.org/view/journals/mwre/139/2/2010mwr3456.1.xml, 2011.

Hamill, T. M., Bates, G. T., Whitaker, J. S., Murray, D. R., Fiorino, M., Galarneau, T. J., Zhu, Y., and Lapenta, W.: NOAA's Second-Generation Global Medium-Range Ensemble Reforecast Dataset, Bull. Amer. Meteor. Soc., 94, 1553–1565, https://doi.org/10.1175/bams-d-12-00014.1, https://doi.org/10.1175/BAMS-D-12-00014.1, 2013.

Islam, K., Rahman, M. F., and Jashimuddin, M.: Modeling land use change using Cellular Automata and Artificial Neural Network: The case of Chunati Wildlife Sanctuary, Bangladesh, Ecological Indicators, 88, 439–453, https://www.sciencedirect.com/science/article/pii/S1470160X18300554, 2018.

Jankov, I., Gallus, W. A., Segal, M., and Koch, S. E.: Influence of Initial Conditions on the WRF?ARW Model QPF Response to Physical Parameterization Changes, Weather and Forecasting, 22, 501–519, https://doi.org/10.1175/WAF998.1, https://journals.ametsoc.org/view/journals/wefo/22/3/waf998_1.xml, 2007.

Jankov, I., Grasso, L. D., Sengupta, M., Neiman, P. J., Zupanski, D., Zupanski, M., Lindsey, D., Hillger, D. W., Birkenheuer, D. L., Brummer, R., and Yuan, H.: An Evaluation of Five ARW-WRF Microphysics Schemes Using Synthetic GOES Imagery for an Atmospheric River Event Affecting the California Coast, Journal of Hydrometeorology, 12, 618–633, https://doi.org/10.1175/2010JHM1282.1, https://journals.ametsoc.org/view/journals/hydr/12/4/2010jhm1282_1.xml, 2011.

Jankov, I., Berner, J., Beck, J., Jiang, H., Olson, J. B., Grell, G., Smirnova, T. G., Benjamin, S. G., and Brown, J. M.: A Performance Comparison between Multiphysics and Stochastic Approaches within a North American RAP Ensemble, Monthly Weather Review, 145, 1161–1179, https://doi.org/10.1175/MWR-D-16-0160.1, https://journals.ametsoc.org/view/journals/mwre/145/4/mwr-d-16-0160.1.xml, 2017.

Jankov, I., Beck, J., Wolff, J., Harrold, M., Olson, J. B., Smirnova, T., Alexander, C., and Berner, J.: Stochastically Perturbed Parameterizations in an HRRR-Based Ensemble, Mon. Wea. Rev., 147, 153–173, https://doi.org/10.1175/mwr-d-18-0092.1, https://doi.org/10.1175/MWR-D-18-0092.1, 2019.

Jin, J., Miller, N. L., and Schlegel, N.: Sensitivity study of four land surface schemes in the WRF model, Advances in Meteorology, 2010, 2010.

Jülich Supercomputing Centre: JUWELS: Modular Tier-0/1 Supercomputer at the Jülich Supercomputing Centre, Journal of large-scale research facilities, 5, https://doi.org/10.17815/jlsrf-5-171, http://dx.doi.org/10.17815/jlsrf-5-171, 2019.

Kleczek, M. A., Steeneveld, G.-J., and Holtslag, A. A. M.: Evaluation of the Weather Research and Forecasting Mesoscale Model for GABLS3: Impact of Boundary-Layer Schemes, Boundary Conditions and Spin-Up, Boundary-Layer Meteorology, 152, 213–243, https://doi.org/10.1007/s10546-014-9925-3, https://doi.org/10.1007/s10546-014-9925-3, 2014.

Knist, S., Goergen, K., Buonomo, E., Christensen, O. B., Colette, A., Cardoso, R. M., Fealy, R., Fernández, J., García-Díez, M., Jacob, D., Kartsios, S., Katragkou, E., Keuler, K., Mayer, S., Meijgaard, E., Nikulin, G., Soares, P. M. M., Sobolowski, S., Szepszo, G., Teichmann, C., Vautard, R., Warrach-Sagi, K., Wulfmeyer, V., and Simmer, C.: Land-atmosphere coupling in EURO-CORDEX evaluation experiments, Journal of Geophysical Research: Atmospheres, 122, 79–103, https://doi.org/10.1002/2016JD025476, https://doi.org/10.1002/2016JD025476, 2017.

Knist, S., Goergen, K., and Simmer, C.: Evaluation and projected changes of precipitation statistics in convection-permitting WRF climate simulations over Central Europe, Climate Dynamics, https://doi.org/10.1007/s00382-018-4147-x, 2018.

Leutbecher, M. and Palmer, T. N.: Ensemble forecasting, Predicting weather, climate and extreme events, 227, 3515–3539, http://www.sciencedirect.com/science/article/pii/S0021999107000812, 2008.

Li, C.-H., Berner, J., Hong, J.-S., Fong, C.-T., and Kuo, Y.-H.: The Taiwan WRF Ensemble Prediction System: Scientific Description, Model-Error Representation and Performance Results, Asia-Pacific Journal of Atmospheric Sciences, https://doi.org/10.1007/s13143-019-00127-8, 2019.

Liu, X., He, J., Yao, Y., Zhang, J., Liang, H., Wang, H., and Hong, Y.: Classifying urban land use by integrating remote sensing and social media data, null, 31, 1675–1696, https://doi.org/10.1080/13658816.2017.1324976, https://doi.org/10.1080/13658816.2017.1324976, 2017.

Lu, Z., Zhang, Q., Miller, P. A., Zhang, Q., Berntell, E., and Smith, B.: Impacts of Large-Scale Sahara Solar Farms on Global Climate and Vegetation Cover, Geophys Res Lett, 48, e2020GL090 789, https://doi.org/10.1029/2020GL090789, https://doi.org/10.1029/2020GL090789, 2021.

Mooney, P. A., Mulligan, F. J., and Fealy, R.: Evaluation of the Sensitivity of the Weather Research and Forecasting Model to Parameterization Schemes for Regional Climates of Europe over the Period 1990-95, Journal of Climate, 26, 1002–1017, https://doi.org/10.1175/JCLI-D-11-00676.1, https://journals.ametsoc.org/view/journals/clim/26/3/jcli-d-11-00676.1.xml, 2013.

Neumann, P., Düben, P., Adamidis, P., Bauer, P., Brück, M., Kornblueh, L., Klocke, D., Stevens, B., Wedi, N., and Biercamp, J.: Assessing the scales in numerical weather and climate predictions: will exascale be the rescue?, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 377, 20180 148, https://doi.org/10.1098/rsta.2018.0148, https://doi.org/10.1098/rsta.2018.0148, 2019.

Otkin, J. A. and Greenwald, T. J.: Comparison of WRF Model-Simulated and MODIS-Derived Cloud Data, Mon. Wea. Rev., 136, 1957–1970, https://doi.org/10.1175/2007mwr2293.1, https://doi.org/10.1175/2007MWR2293.1, 2008.

Palmer, T. N.: Predicting uncertainty in forecasts of weather and climate, Reports on Progress in Physics, 63, 71–116, http://dx.doi.org/10.1088/0034-4885/63/2/201, 2000.

Palmer, T. N.: Towards the probabilistic Earth-system simulator: a vision for the future of climate and weather prediction, Q.J.R. Meteorol. Soc., 138, 841–861, https://doi.org/10.1002/qj.1923, https://doi.org/10.1002/qj.1923, 2012.

Palmer, T. N. and Hardaker, P. J.: Handling uncertainty in science, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 369, 4681–4684, https://doi.org/10.1098/rsta.2011.0280, https://doi.org/10.1098/rsta.2011.0280, 2011.

Pieri, A. B., von Hardenberg, J., Parodi, A., and Provenzale, A.: Sensitivity of Precipitation Statistics to Resolution, Microphysics, and Convective Parameterization: A Case Study with the High-Resolution WRF Climate Model over Europe, Journal of Hydrometeorology, 16, 1857–1872, https://doi.org/10.1175/JHM-D-14-0221.1, https://journals.ametsoc.org/view/journals/hydr/16/4/jhm-d-14-0221_1.xml, 2015.

Pontius, R.: Quantification error versus location error in comparison of categorical maps (vol 66, pg 1011, 2000), Photogrammetric Engineering and Remote Sensing, 67, 540–540, 2001.

Rögnvaldsson, O., Bao, J.-W., Àgùstsson, H., and Òlafsson, H.: Downslope windstorm in Iceland WRFMM5 model comparison, Atmospheric Chemistry and Physics, 11, 103–120, https://doi.org/10.5194/acp-11-103-2011, https://acp.copernicus.org/articles/11/103/2011/, 2011.

Rohrig, K., Berkhout, V., Callies, D., Durstewitz, M., Faulstich, S., Hahn, B., Jung, M., Pauscher, L., Seibel, A., Shan, M., Siefert, M., Steffen, J., Collmann, M., Czichon, S., Dörenkämper, M., Gottschall, J., Lange, B., Ruhle, A., Sayer, F., Stoevesandt, B., and Wenske, J.: Powering the 21st century by wind energy–Options, facts, figures, Applied Physics Reviews, 6, 031 303, https://doi.org/10.1063/1.5089877, https://doi.org/10.1063/1.5089877, 2019.

Samaniego, L., Thober, S., Kumar, R., Wanders, N., Rakovec, O., Pan, M., Zink, M., Sheffield, J., Wood, E. F., and Marx, A.: Anthropogenic warming exacerbates European soil moisture droughts, Nature Climate Change, 8, 421–426, https://doi.org/10.1038/s41558-018-0138-5, https://doi.org/10.1038/s41558-018-0138-5, 2018.

Santos-Alamillos, F. J., Pozo-Vázquez, D., Ruiz-Arias, J. A., Lara-Fanego, V., and Tovar-Pescador, J.: Analysis of WRF Model Wind Estimate Sensitivity to Physics Parameterization Choice and Terrain Representation in Andalusia (Southern Spain), J. Appl. Meteor. Climatol., 52, 1592–1609, https://doi.org/10.1175/jamc-d-12-0204.1, https://doi.org/10.1175/JAMC-D-12-0204.1, 2013.

Schneider, L. C. and Gil Pontius, R.: Modeling land-use change in the Ipswich watershed, Massachusetts, USA, Agriculture, Ecosystems & Environment, 85, 83–94, https://doi.org/10.1016/S0167-8809(01)00189-X, https://www.sciencedirect.com/science/article/pii/S016788090100189X, 2001.

Sillmann, J., Thorarinsdottir, T., Keenlyside, N., Schaller, N., Alexander, L. V., Hegerl, G., Seneviratne, S. I., Vautard, R., Zhang, X., and Zwiers, F. W.: Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities, Weather and Climate Extremes, 18, 65–74, https://doi.org/10.1016/j.wace.2017.10.003, https://www.sciencedirect.com/science/article/pii/S2212094717300440, 2017.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D., Duda, M. G., H., X.-Y., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 3, Tech. Rep. NCAR/TN-475+STR, University Corporation for Atmospheric Research, https://doi.org/doi:10.5065/D68S4MVH, 2008.

Stengel, M., Kniffka, A., Meirink, J. F., Lockhoff, M., Tan, J., and Hollmann, R.: CLAAS: the CM SAF cloud property data set using SEVIRI, Atmos. Chem. Phys., 14, 4297–4311, https://acp.copernicus.org/articles/14/4297/2014/, 2014.

Stergiou, I., Tagaris, E., and Sotiropoulou, R.-E. P.: Sensitivity Assessment of WRF Parameterizations over Europe, Proceedings, 1, https://doi.org/10.3390/ecas2017-04138, 2017.

Stöckli, R., Duguay–Tetzlaff, A., Bojanowski, J., Hollmann, R., Fuchs, P., and Werscheck, M.: CM SAF ClOud Fractional Cover dataset from METeosat First and Second Generation - Edition 1 (COMET Ed. 1), https://doi.org/10.5676/EUM_SAF_CM/CFC_METEOSAT/V001, https://doi.org/10.5676/EUM_SAF_CM/CFC_METEOSAT/V001, 2017.

Stöckli, R., Bojanowski, J. S., John, V. O., Duguay-Tetzlaff, A., Bourgeois, Q., Schulz, J., and Hollmann, R.: Cloud Detection with Historical Geostationary Satellite Sensors for Climate Applications, Remote Sensing, 11, https://doi.org/10.3390/rs11091052, https://www.mdpi.com/2072-4292/11/9/1052, 2019.

Tabari, H.: Climate change impact on flood and extreme precipitation increases with water availability, Scientific Reports, 10, 13 768, https://doi.org/10.1038/s41598-020-70816-2, https://doi.org/10.1038/s41598-020-70816-2, 2020.

Tomaszewski, J. M. and Lundquist, J. K.: Simulated wind farm wake sensitivity to configuration choices in the Weather Research and Forecasting model version 3.8.1, Geosci. Model Dev., 13, 2645–2662, https://gmd.copernicus.org/articles/13/2645/2020/, 2020.

van Leeuwen and Jan, P.: Particle Filtering in Geophysical Systems, Mon. Wea. Rev., 137, 4089–4114, https://doi.org/10.1175/2009mwr2835.1, https://doi.org/10.1175/2009MWR2835.1, 2009.

Varga, A. J. and Breuer, H.: Sensitivity of simulated temperature, precipitation, and global radiation to different WRF configurations over the Carpathian Basin for regional climate applications, Climate Dynamics, 55, 2849–2866, https://doi.org/10.1007/s00382-020-05416-x, https://doi.org/10.1007/s00382-020-05416-x, 2020.

Varga, G.: Changing nature of Saharan dust deposition in the Carpathian Basin (Central Europe): 40 years of identified North African dust events (1979-2018), Environment International, 139, 105 712, http://www.sciencedirect.com/science/article/pii/S0160412019349955, 2020.

Warrach-Sagi, K., Schwitalla, T., Wulfmeyer, V., and Bauer, H.-S.: Evaluation of a climate simulation in Europe based on the WRF-NOAH model system: precipitation in Germany, Climate Dynamics, 41, 755–774, https://doi.org/10.1007/s00382-013-1727-7, 2013.

Yan, L., Eugenia, K., Safa, M., Jorge, R., Fred, K., Daniel, K.-D., Eviatar, B., and Ning, Z.: Climate model shows large-scale wind and solar farms in the Sahara increase rain and vegetation, Science, 361, 1019–1022, https://doi.org/10.1126/science.aar5629, https://doi.org/10.1126/science.aar5629, 2018.

Yuan, F., Sawaya, K. E., Loeffelholz, B. C., and Bauer, M. E.: Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing, Remote Sensing of Environment, 98, 317–328, https://doi.org/10.1016/j.rse.2005.08.006, https://www.sciencedirect.com/science/article/pii/S0034425705002646, 2005.