Sensitivity of Convective Forecasts to Driving and Regional Models During the 2020 Hazardous Weather Testbed

November 2021

Forecasting Research

Technical Report No: 649

Written by Regional Systems Evaluation Team, RMED, Met Office:

David L. A. Flack, Caroline Bain, James Warner

Collaborators:
Adam Clark, Burkely Gallo, Israel Jirak, Brett Roberts, Larissa Reames: NOAA/ NSSL
Craig Schwartz: NCAR

© Crown Copyright 2021, Met Office

© Crown copyright 2021, Met Office

Met Office

Contents

© Crown copyright 2021, Met Office

Executive Summary

Each year during the peak severe convective storm season in the USA (early spring) NOAA's SPC and NSSL run a Hazardous Weather Testbed Spring Forecasting Experiment (HWT SFE). The HWT aims to bring together operational meteorologists, research scientists and academics from across the USA, and the globe, to consider several experiments with a focus on convective-scale modelling. The HWT is a real time tool to investigate different scientific questions that have practical use for forecasting of severe convection. In addition it acts to enhance R2O and O2R activities, and build relations between different national weather centres.

During the HWT SFE 2020 the Met Office contributed two experiments: i) an ensemble experiment and ii) a deterministic experiment focusing on the impact of driving models on regional models. The latter experiment is discussed here. Regional models are traditionally run for limited areas across the globe. In running over limited areas they require lateral boundary conditions and initial conditions that are, often, from a global model (the driving model). The aspects of severe convection that are sensitive to the driving model or the regional model remain unclear.

An experiment in HWT SFE 2020 compared three regional models (WRF, FV3, UM) that were 'driven' by two different global models (GFS and UM). Through this experiment new technical capability was developed to run UM regional model driven from GFS initial conditions and it was also the first time the FV3 regional model was driven from UM Global Model. This experiment also showed that the Met Office were able to successfully transfer the UM Global model 00 UTC forecast files to USA in a timely enough manner for WRF and FV3 to run (and produce plots) in time for the daily HWT Evaluation Discussions at 17 UTC.

In terms of results, there was some (weak) subjective indication that in strong large-scale forced events driving models dominated the forecasts whereas in weak large-scale conditions the regional model cores were more influential. There is some quantitative evidence (using new convection diagnostics) that the convective structure is primary sensitive to the regional model in terms of how fragmented storm cells are and the ratio of convective to stratiform precipitation within convective events.

Initial analysis has indicated that there is model sensitivity to experimental set up, especially in the initial part of the forecast, and this has meant it is not possible to make definitive conclusions on the impact of the driving model on convection. Sensitivities to the different model setups are examined and it is shown that there is a noticeable impact of changing the driving soil state on results.

Recommendations are made on the future setup of this type of experiment, given the findings from HWT SFE 2020. These recommendations include the use of regional model native soil state regardless of driving model; the same domain for all models; and the same driving data. Improving the comparisons will enable the detection of whether errors in the model are coming from the regional model (core or parametrizations) or from the driving model (initiation conditions, boundary conditions, global data assimilation), and so can then indicate areas where the model can be improved on both regional and global scales.

© Crown copyright 2021, Met Office

1. Introduction

Every year beginning in late April or early May (to coincide with the peak convective storm activity across the Great Plains of the USA) NOAA's Storm Prediction Centre, in conjunction with their National Severe Storms Laboratory, have a five-week intensive testbed: the Hazardous Weather Testbed Spring Forecasting Experiment (HWT SFE). In 2020 due to the COVID-19 pandemic this was changed from an in-person testbed to a, reduced, virtual testbed.

As in recent years the Met Office contributed to the HWT SFE. This year the Met Office provided model output for two experiments: i) an ensemble experiment investigating time-lagging and the impact of small multi-model ensembles which were evaluated by a new subjective scoring technique developed by Nigel Roberts; and ii) a deterministic sensitivity experiment to determine the impact of driving models *vs.* regional models on the forecasts of severe convection.

The ensemble experiment was led by Aurore Porson and a separate report has been written on those activities (Porson et al. 2020). The focus of this report is the deterministic experiment: driving model *vs*. regional model sensitivities.

The deterministic experiment arose out of discussions during a Convection Working Group workshop in January 2020. The main premise of the deterministic experiment is to discover what aspects of severe convection forecasts are controlled by the large-scale driving model, and which aspects come from the regional model (convection-permitting configurations). The experiment also aims at determining how the relative importance of the driving and regional model evolves throughout the forecast. The idea of the experiment is to show where improvements can be made (or the current limitations are) for forecasts of convection: the model core/relevant parametrizations or from the boundary conditions or initial conditions/global data assimilation. In terms of the Met Office forecasting system this will have the greatest benefit for the regional models that do not have their own analysis (i.e. all regional models except the UKV) and instead get their initial conditions directly from the global model.

This type of experiment has been used on many different scales and is, perhaps, more frequently used on the climate scale, where it was originally suggested by Phillips et al. (2004) and culminated in the so-called `Transpose-AMIP' experiments (e.g. Williams et al. 2013; Ma et al. 2013; Bony et al. 2013; Roff 2015; Pearson et al. 2015; Li et al. 2018; Sexton et al. 2019; Brient et al. 2019; Flack et al. 2021b). However, it is not just at the climate scale where these types of experiments have been performed, they have been [or are currently

© Crown copyright 2021, Met Office

underway in the case of a global model comparison initiated from ECMWF analyses (Duncan Ackerley: personal communication 2020)] on mesoscale models (grid length 25 km) as part of the precursor to the Short Range Numerical Weather Prediction – Ensemble Prediction Systems (SRNWPEPS) programme (Garcia-Moya et al. 2011) and also at 7 km grid lengths (Marglisi et al. 2014). Investigations have occurred along similar lines for convection-permitting models as well. These convection-permitting model experiments are usually in the context of ensembles as opposed to deterministic models (e.g. Keil et al. 2014; Kühnlein et al. 2014; Porson et al. 2019).

Many of these experiments focus on the impact of one driving model and less of an impact on comparisons –-- so the focus is on model biases or representation of physical processes. However, where multiple models and/or multiple driving conditions are used, model comparisons to help identify and determine the source or cause of the model error are more common than verification studies (e.g. Williams et al. 2013, Porson et al. 2019). Verification from this style of experiment occasionally occurs. In Garcia-Moya et al. 2011, results indicated that in a mid-latitude domain, the five different regional models tested (COSMO, HIRLAM, HRM, MM5 and UM) performed best with their native/ normal driving model (either GFS, GME, IFS or UM). Interest in verification becomes more acute perhaps when models are used that do not have a native driving model –-- i.e. the model does not produce its own analysis (e.g. LMDZ or WRF) or in the tropics where some global models perform more strongly than others. For example, in Porson et al. 2019 it was shown that the regional UM performed best with IFS driving model (than UM global) over a domain covering Peninsula Malaysia and this was in part attributed to a better IFS analysis.

The published works only partially document the literature on this subject; there are many conference presentations of this work (e.g. Karmalkar 2015; Roff and Zhang 2015; Karmalkar and Bradley 2016; Burkhardt et al. 2017). The methodology for running these experiments pays close attention to ensuring as clean a comparison as possible between a control (e.g. regional model run from its native driving model) and an experiment (e.g. regional model run from a non-native driving model), with all other elements of influence on the outcome minimised. Detailed descriptions of experimental design are available in the climate experiment papers (e.g. Phillips et al. 2004, Williams et al. 2013; Roff 2015).

The rest of this report is set out as follows: Section 2 considers the experimental setup, models and diagnostics; Section 3 discusses the feedback during HWT SFE 2020 and more objective measures indicating the differences between the driving and regional models; Section 4 considers the caveats and sensitivities on these results; Section 5 indicates the lessons learnt from this experiment and indicates a set of recommendations for this type of

© Crown copyright 2021, Met Office

experiment at the convective scale; Section 6 summarises the report and provides conclusions based on the 2020 results.

2. Experimental Setup

Here we discuss the model configurations used and sensitivity tests examined (Section 2.1) and the diagnostics created for this work alongside the different comparison options (Section 2.2). To take convective variability into account in these sensitivity tests a large number of cases (as opposed to an ensemble) are considered to alleviate concerns raised in Flack et al. (2019) of sensitives being claimed with limited statistical backing due to consideration of single forecasts or at most three forecasts. The work allows statistical analysis to take place.

2.1. Model configurations

In the HWT SFE 2020 deterministic driving vs. regional model sensitivity test three regional models are all nested inside two global models. The regional models used are the FV3, WRF and the UM; the global models used are the GFS and UM. All regional models use convection-permitting configurations (3 km grid lengths for WRF and FV3, 2.2 km grid length for the UM) with their output interpolated onto the 3 km Community Leveraged Unified Ensemble grid (CLUE; Clark et al. 2018) to give as similar output as possible, given different domain sizes, for (more) direct comparisons of the forecasts. The model domains are shown in Fig. 1.

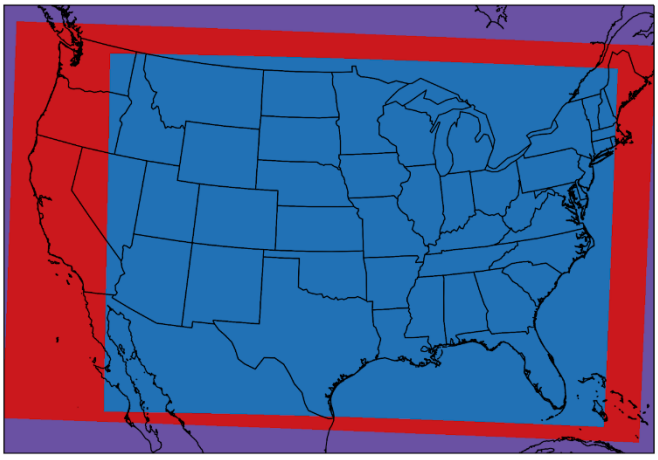


Figure 1: The domains of the different regional models and observations. The observations and FV3 are covered by the purple domain, WRF covers the red domain and the UM the blue domain.

© Crown copyright 2021, Met Office

Table 1 shows the different configurations used for the model setups. The table shows that there are differences in how the soil state is treated in each of the model experiments. The change in soil state diverges from previous literature discussed in Section 1 where soil moisture from native models or the model's climatology is used throughout. This is because soil moisture is not a directly transferrable parameter and is treated differently in different global models. Further differences from previous literature and potential caveats are discussed in more detail in Sections 4 and 5.

To consider the impact of (some) of these differences two additional experiments were setup and ran after the testbed: i) WRF(UM:GFS LBC) in which everything is kept as WRF-UM apart from the LBCs which are now the GFS LBCs; and ii) WRF(UM:GFS SOIL) which remains the same as WRF-UM except that soil moisture and soil temperature now come from the GFS. The latter experiment is closer to a Transpose-AMIP design.

Differences between the models will include their parametrizations (both at a regional and global scale) – no additional tuning was performed in any of the non-native driving conditions forecasts.

A further factor that needs to be considered in all experiments that use non-native driving conditions is the `initial shock' (Klocke and Rodwell 2014). The initial shock arises because of different balances in different models. Klocke and Rodwell (2014) showed that it takes time for an NWP model to adjust to its native attractor from different driving conditions. This means that the short lead-times should be removed from analysis (e.g. Judd et al 2008).

In this experiment, all regional models were initiated from a 'cold start', meaning there was no data assimilation and initiation takes place from a lower resolution global model, meaning there will also be a spin-up period as the regional model creates higher resolution convective structures. Both initial-shock and spin-up factors imply results at the beginning of the forecast are less robust and may not show consistent features to periods later in the forecast, hence early times (less than T+12 hours) will be neglected in the analysis.

Full experimental details can be found in the HWT SFE 2020 operational plan (Clark et al. 2020a).

Forecasts are considered for 32 of the available 40 cases due to lack of data from all models or missing or corrupted files. Thus, the analysis spans 25 April to 29 May 2020 inclusive missing 2, 7, and 17 May.

© Crown copyright 2021, Met Office

Observations from radar and gridded surface-based variable products are used throughout the work to compare the models to.

Table 1: Core experiment setups and IDs, indicating factors that are different between the different model configurations.

Experiment ID	Regional Model	Driving Model	Lateral Boundary Conditions	Soil Moisture	Soil Temperature
WRF(UMgm)	WRF	UMgm	UMgm	UMgm	UMgm
WRF(GFS)	WRF	GFS	GFS	GFS	GFS
FV3(UMgm)	FV3	UMgm	UMgm	UMgm	UMgm
FV3(GFS)	FV3	GFS	GFS	GFS	GFS
UMrm(UMgm)	UM	UMgm	UMgm	UMgm	UMgm
UMrm(GFS)	UM	GFS	GFS	UMgm	GFS

2.2. Diagnostics and available comparisons

Throughout this work several diagnostics have been used, and histograms of the surface variables have been considered. The calculated diagnostics are described here. All diagnostics are calculated over the UM domain to ensure domain consistency (Fig. 1).

There are numerous comparisons given the six experiments that could occur for these experiments. However, many of these comparisons cannot be used to answer the question about where the regional model or driving model dominates. These comparisons are shown in Fig. 2 and take the form of an upper-triangular matrix. The main comparisons for this work that make the most physical sense are in red and blue in Fig. 2. The blue and red comparisons are used for the different forecasts in the temperature variance and fraction of common points calculations.

A threshold of 30 dBZ is used to identify convective precipitation for all diagnostics. This threshold is lower than traditionally used for severe convection (approximately 40 dBZ) however results remain qualitatively consistent when thresholds are taken higher or lower. The model values are not bias corrected in any of the comparisons, although bias correction could be considered in future work with a quantile-mapping approach.

© Crown copyright 2021, Met Office

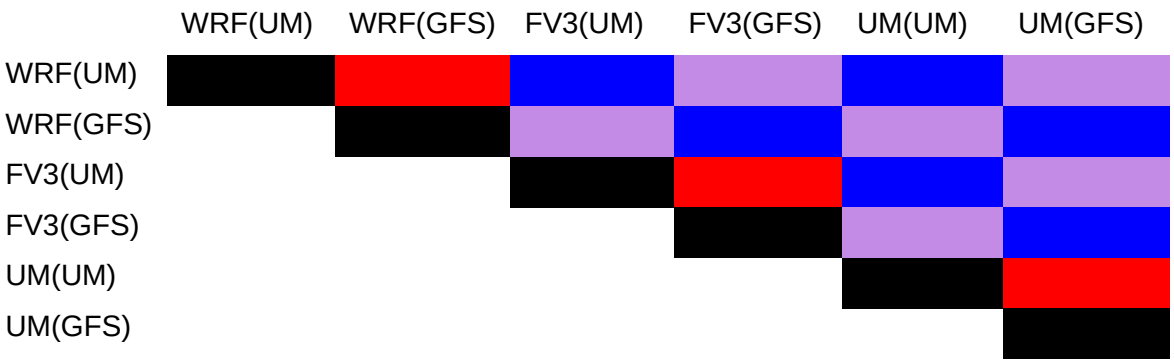


Figure 2: Available comparisons between the models. Black squares represent one-to-one comparisons, red squares represent driving model comparisons, blue squares the regional model comparisons and purple squares comparisons that make less sense as they form a mixture of regional and driving model comparisons.

Temperature variance (DTE_T)

The temperature variance (DTE_T) is the temperature component of the difference total energy (DTE; e.g. Zhang et al. 2003), which is frequently used to consider forecast differences (e.g. Selz and Craig 2015), and has been used in temperature variance form by Flack et al. (2021a). The temperature variance is calculated as

$$\mathrm{DTE_T} = \frac{c_p}{T_{ref}} T'T',$$

for T_{ref}, a reference temperature of 273 K, c_p, the specific heat capacity at constant pressure and T', the difference between the temperature of one forecast compared to another (e.g. Fig. 2).

The DTE_T is used to help determine the mechanisms for the differences in the forecasts. For this study the temperature considered is the 2 m temperature. Furthermore, the DTE_T is computed as a domain average across the UM domain.

Fraction of common points (F_{common})

The Fraction of Common Points (F_{common}) is a simple diagnostic first used in Leoncini et al. (2010), and more recently adapted by Flack et al. (2018) to ensure that F_{common} varies between zero and unity. Flack et al.'s (2018) formulation is used here:

$$\mathrm{F_{common}} = \frac{N_{1,2}}{N_1 + N_2 - N_{1,2}},$$

© Crown copyright 2021, Met Office

for *N*, the number of precipitating points, and the subscripts refer to the forecasts being compared (e.g. Fig. 2), and a subscript (1,2) refers to the points in the same location in both forecasts (i.e. the common points). These comparisons have been made in Flack et al. (2018), Clark et al. (2021) and Flack et al. (2021a) and they show useful concepts for ensemble forecast spread. In the context of this work F_{common} has a marginally different meaning and as such requires the calculation of F_{common} due to both regional and driving model differences for a sensible interpretation to be made. For example, if forecasts have a greater F_{common} between regional model comparisons than driving model comparisons it implies that the driving model dominates in the positioning of the convective events (convective threshold of 30 dBZ), and vice versa. It is worth noting that F_{common} is just as valid with the use of an absolute threshold as well as a percentile threshold. The percentile threshold changes the meaning a little bit based on potential biases and magnitude mis-matches. Here, an absolute threshold is used but this can lead to discrepancies in terms of number of points that meet the threshold between the models used. This has limited impact on the interpretation as a smaller number of points in one model will lead to a lower F_{common} and it will never equal one as points can never fully agree. This means the interpretation becomes either different placement or lack of points. To give an idea of which could be dominating the number of points reaching the threshold in each forecast can be examined.

Convective Fragmentation Index (CFI)

The Convective Fragmentation Index (CFI) was specifically created for the purposes of these comparisons, and to add further information than traditional cell statistics (e.g. Hanley et al. 2014, Stein et al. 2015, Clark et al. 2021). The reason to not focus on traditional cell statistics was because these often focus on size distributions and size is not the only factor that implies how fragmented convection appears in the model. Therefore, inspiration was taken from other fields including ecology (e.g. the landscape dissection index, Bowen and Burgess 1981). The inspiration was used to create a single index in which value can be gained from the variations of this index with either threshold or time but also in comparisons between models and reality. The CFI is calculated as follows

$$\mathrm{CFI} = \frac{1}{N}\sum_{i=0}^{N}\frac{\pi(D_i)^2}{4A_c},$$

for *N,* the number of convective cells (that have been identified through image processing), D_i, the equivalent diameter of each cell (*i*), and A_c, the total area covered by all convective cells. The CFI is somewhat like the 2D shape index used in Pscheidt et al. (2019); the CFI

© Crown copyright 2021, Met Office

focusses on the area of the cells to determine their fragmentation, as opposed to their perimeter.

In its current form, the CFI thus considers two key characteristics of a fragmented field: the size and number of objects. Figure 3 shows idealised examples to aid interpretation of the CFI --- the smaller the CFI the more fragmented the precipitation field is. Figure 3 also indicates that if the CFI is considered over multiple thresholds a “critical” threshold starts to appear in which the convection is naturally organised. Considering the smaller threshold (blue), in Fig. 3, the CFI increases in magnitude following the order Fig. 3a, b, c, d. In Fig. 3a the CFI is close to zero, whereas in Fig. 3d the CFI is equal to one. On the other hand, considering the larger threshold (yellow) natural organisation appears to be occurring with Figs. 3a, b and c all having a similar CFI, despite the differences at the smaller threshold. Furthermore, at this larger threshold Figs. 3a, b and c are not too dissimilar from Fig. 3d (which still has a CFI of one).

The work represented here shows an initial formulation of the CFI. Work is ongoing to improve the CFI to gain a clearer picture of the fragmented field. Therefore, this report acts as a first documentation of the index.

Convective Proportion (CP)

The Convective Proportion (CP) is a simple diagnostic that is applied to each identified convective cell and compares the ratio of convective points to stratiform points:

$$\mathrm{CP} = \frac{N_{convective}}{N_{precipitating}},$$

for $N_{convective}$, the number of convective points (convective threshold of 30 dBZ) and $N_{precipitating}$, the number of points considered to be precipitating (a threshold of 5 dBZ). The CP indicates biases with stratiform rainfall so can be particularly useful for larger events with extensive stratiform regions (such as Mesoscale Convective Systems).

© Crown copyright 2021, Met Office

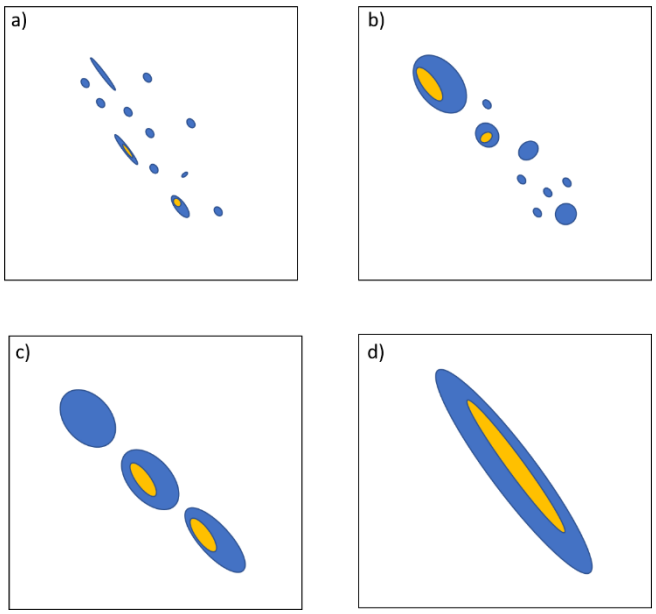


Figure 3: An idealised example to indicate the interpretation of the convective fragmentation index (CFI). Blues indicate a smaller threshold than the yellow areas.

3. Driving *vs.* Regional Model Sensitivities

Two types of analysis have been undertaken with this experiment: subjective evaluation (Section 3.1) and objective evaluation (Section 3.2). Various aspects of the forecasts are examined to determine the evolution of the relative importance of driving and regional models to the forecasts of convection. Furthermore, different aspects of the convective forecast (location, structure, and fragmentation) are examined to determine which model has a greater influence: the regional or driving model.

3.1. Subjective analysis

During the HWT SFE, to aid in the comparison between forecasts the HWT web visualiser was setup to create a 3x3 display, an example of which is shown in Fig. 4. In this display the columns (excluding the observations) represent forecasts with the same driving model and the rows show forecasts with the same regional model. Thus, comparisons across the rows represent driving model differences and comparisons down the columns represent regional model differences.

The subjective assessment took place during the HWT SFE 2020 and analysis was performed by NOAA SPC/NSSL (Clark et al. 2020b). The highlights of their work showed a

© Crown copyright 2021, Met Office

result suggesting a lack of clear preference, throughout the experiment, when asked “did you seem more differences between <*variable*> from models with the same driving model and different regional model, or between from models with the same regional model and different driving model?” The results for reflectivity and updraft helicity produced a mean result of 47.4 and the environmental variables (temperature, dewpoint and CAPE) a mean result of 51.377 where values greater than 50 imply driving model differences are larger and values less than 50 imply the regional model differences are larger. It is worth noting that daily variations were higher and there was strong case-to-case variability, so these headline figures do not show the full picture.

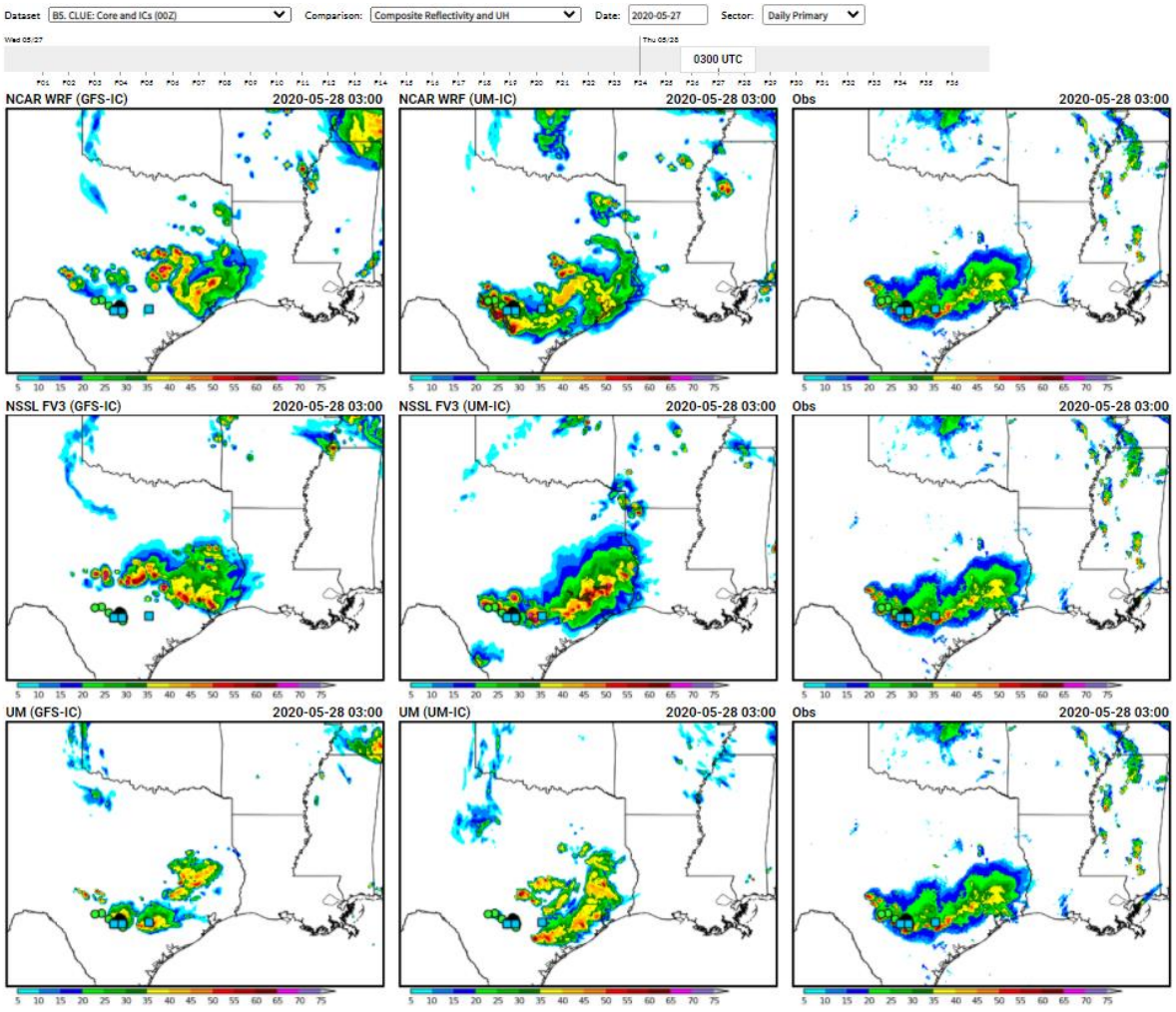


Figure 4: Web viewer display for the driving model *vs*. regional model sensitivity tests during the HWT SFE 2020. Web viewer image courtesy of Brett Roberts/NOAA and available online at: https://hwt.nssl.noaa.gov/sfe_viewer/2020/model_comparisons/?dataset=det_coreic&comparison=cref_uh&sector=hwt_dd1&date=20200529&daily_time=0000&sector_date_offset=0&sector_date=20200529&sector_moving=true.

© Crown copyright 2021, Met Office

In the rest of this section key points noted during the discussions are highlighted, including model biases detected in the UM and subjective analysis performed at the Met Office.

Discussion points during the HWT SFE 2020

Throughout the first three weeks of the HWT SFE 2020 there was not a complete set of model outputs for this experiment. On the other hand, for the last two weeks a full set of model outputs were available for subjective assessment. Thus, the synopsis presented here focuses on these last two weeks. However, it is worth noting that many of the comments occurred in the earlier weeks as well.

The subjective discussions noted a few potential problems in the runs (which led to a change in UM(GFS) simulations for the final week) and were aimed to be fixed or sensitivity to certain factors tested for post-HWT SFE evaluation. These factors were a strange positioning of the lakes in the FV3 forecasts (dipoles in lakes in Wisconsin: Fig. 5a) and an odd behaviour in the UM(GFS) forecasts in which the errors would grow, saturate the colour scale, then shrink and regain spatial structure (Figs. 5b,c,d). The behaviour in the UM(GFS) forecasts is thought to be linked to surface processes or soil state initiation.

Another factor discussed was the, apparent, lack of trend in the stronger sensitivity to the driving model or regional model. However, this often appeared to be based on what the participants were focusing on (i.e. convective mode, structure, positioning or intensity) and thus the interpretation of the wider question presented to the participants may not have fully captured the variation seen. It is also worth noting that some participants noted that several days there was an equal mix of both models influencing the forecasts, whereas others there was a very strong signal for driving model domination (or regional model domination). It is hypothesised that this could be linked to the forcing regime (weak *vs*. strong synoptic/upper-level forcing) but was agreed that there were unlikely to be enough days considered to produce reliable statistics to answer test this hypothesis.

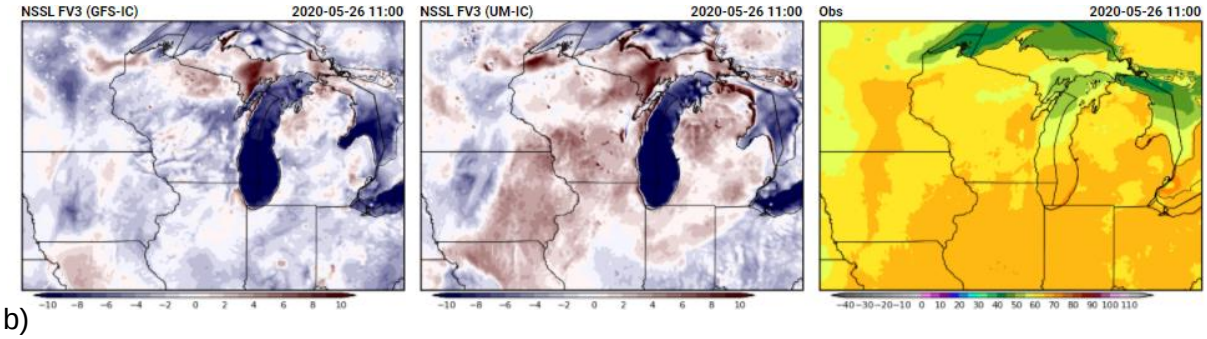
Model biases detected in the UM

Several biases were detected that appeared to be unique to the regional models rather than from the driving model, and this helped lead to the ability to identify distinctive regional model performance throughout the experiment. Here we provide a list of the biases detected within the UM (but also indicate if these biases occurred in the other regional models). The model biases are

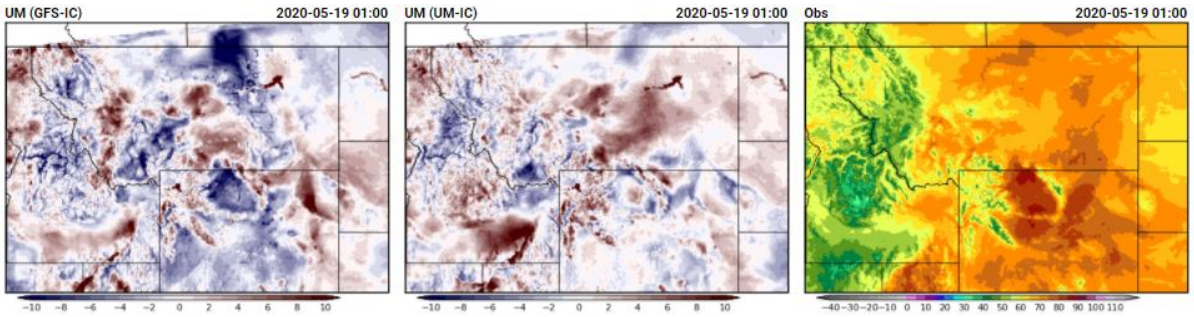
© Crown copyright 2021, Met Office

- a poor diurnal cycle of convection with the UM having the peak too early;
- delayed convection by 1-3 h (occurs in UM, FV3 and WRF);

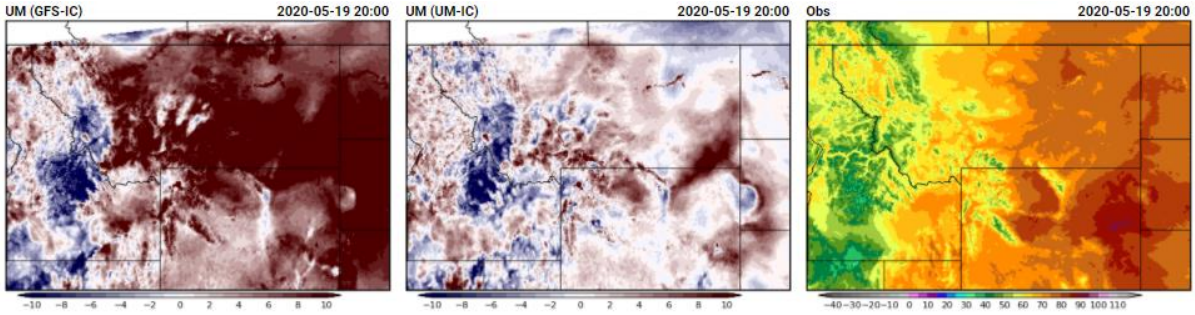
a)



b)



c)



d)

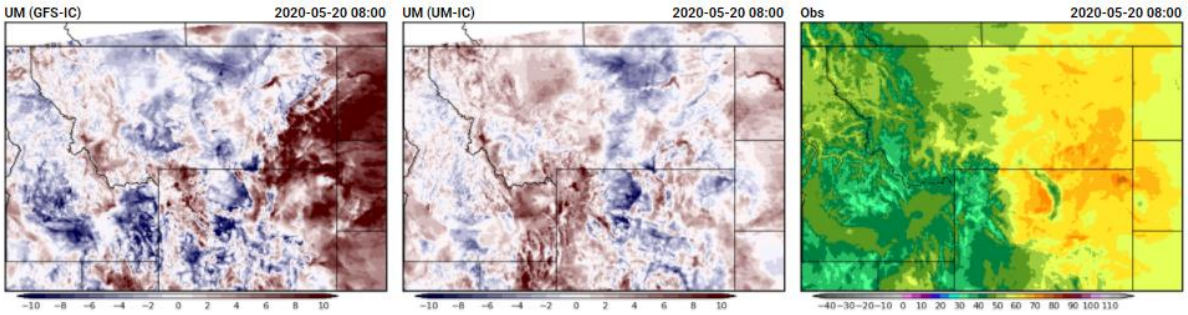


Figure 5: Web viewer showing forecast - observations for 2m temperature a) FV3 simulations at T+11 26 May 2020, and UM simulations at b) T+1 19 May 2020, c) T+20 19 May 2020, d) T+32 from 19 May 2020. Web viewer image courtesy of Brett Roberts/NOAA (link to images in caption of Fig. 4).

© Crown copyright 2021, Met Office

- poor cold pool representation but not in a specific direction (occurs in UM, FV3 and WRF);
- convection regularly decays too early so the lifecycle is not captured;
- the convection is strongly fragmented and at times ‘blobby’ in character;
- there is often too little stratiform rain in large organised mesoscale convective system like events;
- the dewpoint temperatures are too dry (occurs in UM, FV3 and WRF);
- the 2 m temperatures are too warm;
- elevated convection is often missed in the model (occurs in UM, FV3 and WRF);
- lack of upscale growth of convection.

Many of these biases are well-known in the UM and are being investigated in other work.

Met Office subjective analysis

A subjective (internal) questionnaire at the Met Office asked different details to the NOAA questionnaires. However, due to the intense nature of the HWT, the Met Office questionnaire only received 16 responses from a limited selection of people. To put this into context the HWT subjective analysis is based on all participants for all cases, so 20 per day across 5 weeks is approximately 500 responses. Despite the small number of responses, the results are representative of what was seen during the HWT and match reasonably well with NOAA subjective analysis. Figure 6 shows the key questions looking at intensity, timing, structure and location of convection and the relative dominance of the regional and driving model. The available answers to all the questions relating to the relative differences between regional and driving models were i) greater between regional models with same driving conditions; ii) greater between a single regional model with different driving conditions; iii) Neutral; and iv) No difference. The neutral option refers to cases where there are differences in the simulations from driving model and regional model, but it is harder to tell which dominates.

Figure 6 shows that for timing and location of convection the driving model (and as such specification of the large-scale conditions) dominates. On the other hand, the convective structure is possibly more associated with the regional model (although this is case dependent). Differences in the maximum intensity appear to be small and, if they do exist, hard to determine which model appears to be dominating. The case dependence is

© Crown copyright 2021, Met Office

thought to be linked to large-scale forcing however due to data constraints this is not tested. The subjective differences noted throughout both Met Office and NOAA subjective analysis and the discussions during the HWT SFE indicate that further analysis into the structure of convection could be required. It is also indicative of factors that do not usually appear in more 'traditional' objective analysis and so the diagnostics discussed in Section 2.2 that focus on objective-oriented measures could be particularly useful in determining the differences between models and the impact of driving *vs*. regional models for the forecasts of convection.

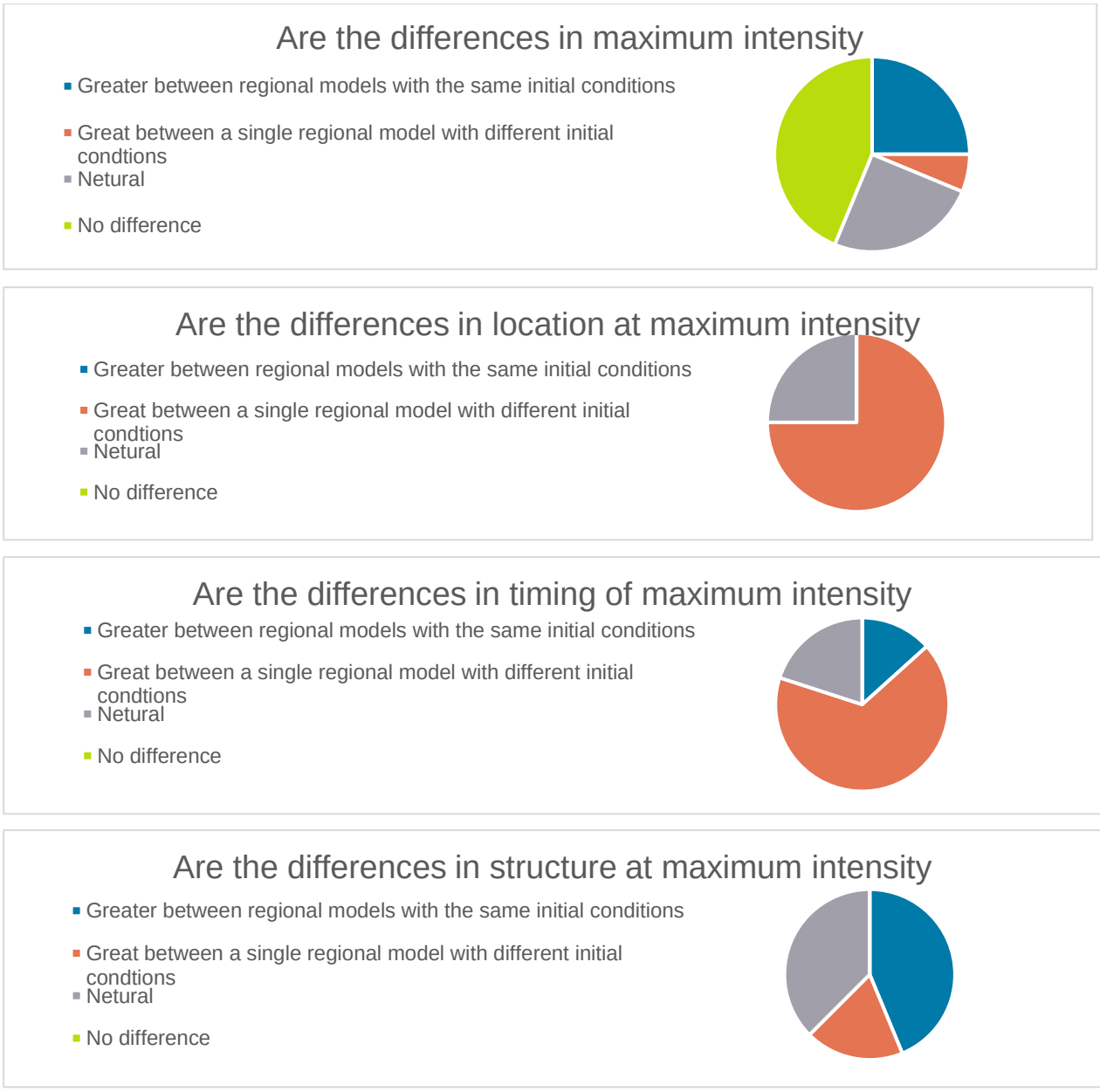


Figure 6: Responses to four of the questions asked in the Met Office questionnaire on the dominance of driving models or regional models during HWT 2020.

© Crown copyright 2021, Met Office

3.2. Objective analysis

Initial shock and DTE_T

As previously discussed, initial shock and spin-up could arise in these simulations, and one such tool for identifying these factors is the DTE_T (Fig. 7). A DTE_T equivalent to c_p/T_{ref} ≈ 3.7 J kg^{-1} is equivalent to a forecast difference of 1 K. The DTE_T's evolution can give clues as to the factors leading to the differences between the forecasts. The DTE_T would increase as errors grow and decrease when errors are reducing. Therefore, the DTE_T is expected to grow from the start of the forecast, grow more rapidly when convection is present and begin to decay as convection dissipates (e.g. Flack et al. 2021a).

To identify the presence of initial shock the driving model DTE_T comparisons are examined (Fig. 7a). Each of the models show some common factors but are also somewhat different to one another as well. The first common factor is an initial rise lasting from T+1—T+3 (although this is reduced in WRF compared to the UM and FV3). This is associated with spin-up as the convective-scale differences start to occur. However, in all models there is a decrease of varying speeds (FV3 is the slowest, then WRF and then the UM shows the fastest drop). This is not expected behaviour for this diagnostic and lasts until T+16 in all models. Whilst some of this drop may be due to decaying convection from the previous day (local time) most of this behaviour is associated with the adjustment from using non-native initial conditions. The adjustment happens to ensure the model returns to its own attractor and that balances within the model are kept consistent (each model has its own balance; e.g. Klocke and Rodwell 2014).

On average, the adjustment period lasts 16 h and so, for the most part, all other diagnostics are considered after T+16. After the adjustment, from the initial shock, the DTE_T grows with the diurnal cycle peaking first in UM comparisons as would be expected from the growth of convective activity. The influence of the boundary conditions is noted clearly in UM and WRF after the drop as the convection reduces. It is worth noting that the FV3 driving model comparisons show a behaviour that is like the regional model comparisons (Fig. 7b) rather than the other driving model comparisons (reasons for this are unknown).

The regional model comparisons (Fig. 7b) that use non-native soil moisture all show interesting behaviour during the first 15h regardless of which models (and driving conditions are being compared). This behaviour shows a sharp increase in differences, a plateau (or slow decrease) and then a sharp decrease again (a little like a square wave).

© Crown copyright 2021, Met Office

This behaviour occurs at varying amplitude and is likely due to changes in the soil state (soil moisture) between the runs (Table 1; Section 4). After this period growth appears to be more associated with differences in convection (due to the rise occurring with the increase in convective activity with the diurnal cycle).

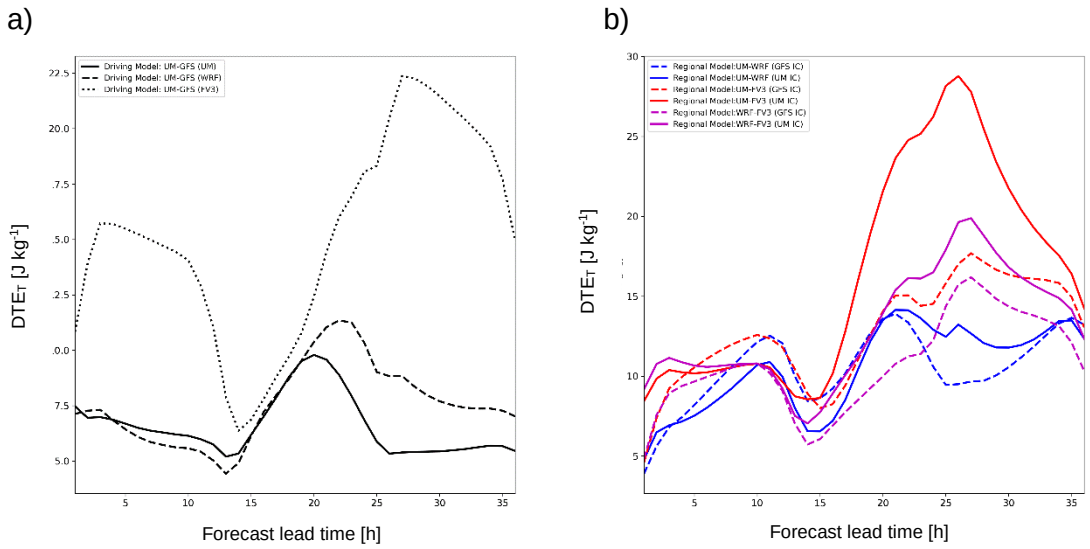


Figure 7: The average temperature variance across the full data set (32 cases) focused on the UM domain a) driving model comparisons and b) regional model comparisons with red lines UM-FV3 comparisons; blue lines UM-WRF comparisons and magenta lines FV3-WRF comparisons.

The comparisons using UM(GFS) forecasts (which use UM soil moisture and GFS soil temperature and hence are more consistent with the TRANSPOSE-AMIP protocol) show the expected behaviour of growth from the start of the forecast that clearly grows and decays with convective activity. The results are sensitive to the soil temperature (Section 4) though this is in part related to the variable considered. Thus UM(GFS) comparisons are likely to be more 'clean' than the FV3(UM) and WRF(UM) comparisons which look to be influenced by the UM soil moisture, and as such these runs should be treated with caution or ideally replaced by runs with native soil moisture.

The impact of the change in soil state means that the question surrounding the evolution of the relative importance of the driving and regional model cannot be fully examined as the soil state contaminates the results, and thus not allow a clean comparison. The impact of soil state on all the results is investigated further in Section 4.

© Crown copyright 2021, Met Office

Considering these results, for the remainder of this section only data after the adjustment period has subsided. Thus, the period considered is T+16 to T+36, unless spin-up is specifically considered useful to consider, such as in the development of the structure of convection.

Distributions of surface variables

One impact that needs to be considered in future is whether there are any changes between distributions of different variables in the models. This might impact the biases and the interpretation of future diagnostics. Therefore, distributions of the surface-based fields examined by participants during the HWT SFE are examined. Figure 9 shows the histograms for four of these fields.

For the most part differences between the 2 m temperature distributions are small (Fig. 9a); larger differences occur for the 2 m dewpoint temperatures, 10 m windspeeds, and the composite radar reflectivity (Figs. 9b,c,d). There appears to be greater differences between regional models than driving models as the distributions associated with the different driving models are more similar than those from different regional models (Fig. 9).

The dewpoint fields show greatest variability of the two temperature distributions (compare Figs. 9a and b) with the greatest difference occurring at the warmer (moister) dewpoints with UM and FV3 forecasts peaking at marginally cooler (drier) temperatures than WRF forecasts. The differences in the dewpoint temperatures will likely be a combination of results from different surface schemes and states, but also differences in the humidity.

The 10 m windspeeds show larger differences (Fig. 9c). The UM simulations, whilst being very similar to the observations from 6 m s^{-1} onwards has a larger occurrence of slower windspeeds than the other models (and observations). On the other hand, both WRF and FV3 are somewhat similar in having more frequent mid-range and faster windspeeds (4—14 m s^{-1}) than the UM or observations suggesting more energetic models or less impact of frictional drag in the boundary layer. Fig. 9c also clearly highlights greater sensitivity to regional model than driving model for windspeed.

© Crown copyright 2021, Met Office

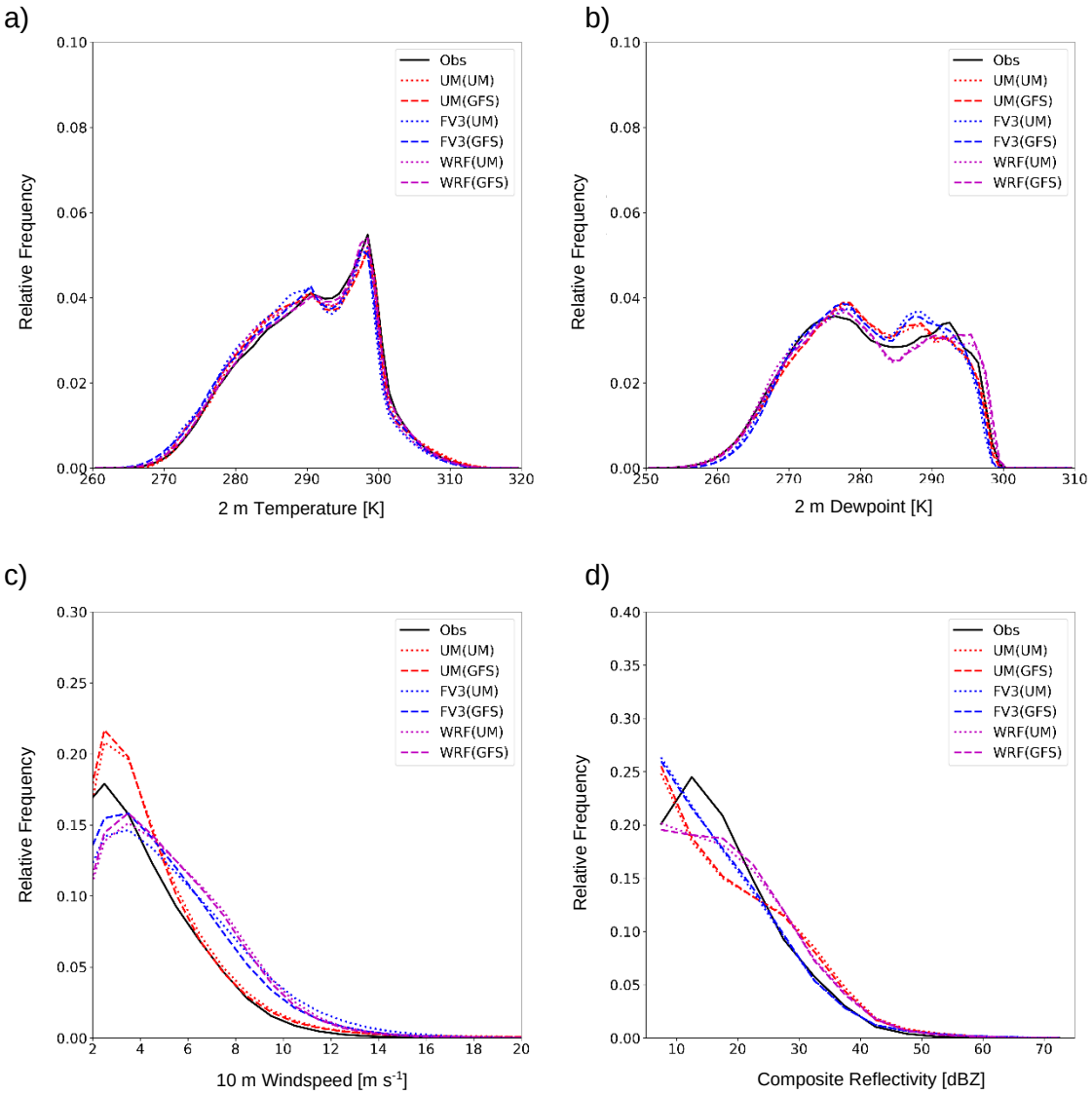


Figure 9: Relative frequency histograms across the UM domain for the whole HWT period after T+16 h for a) 2m temperature, b) 2 m dewpoint temperature, c) 10 m windspeed and d) composite radar reflectivity. Observations are in black, UM regional model is red, FV3 regional model is blue, WRF regional model is magenta, when the UM is the driving model lines are dotted and when the GFS is the driving model lines are dashed.

As with the windspeed the composite radar reflectivity (Fig. 9d) shows some key differences across all models, and stronger sensitivity to regional model than to driving model. The FV3 simulations closely follow the observed reflectivity from 20 dBZ with only a marginal positive bias. For values less than 20 dBZ there is an underestimation of these

© Crown copyright 2021, Met Office

reflectivities (associated with stratiform rain). However, this is not as strong as in other models. Both WRF and UM simulations show a clear positive bias from values greater than 25 dBZ and a clear reduction in stratiform rain, although it is perhaps worth noting that the UM shows greater very low reflectivities (less than 10 dBZ) than WRF. These differences (appearing to be somewhat stronger between regional models than driving models) likely point to the impact of microphysics schemes and the initiation of convection.

It is worth noting that the differences between the simulations (for any comparisons, regional model for driving model) are not statistically significant at the 95% confidence interval when using a Wilcoxon Rank-Signed Test (the most appropriate statistical test for comparing these distributions; e.g. Wilks, 2011). This is particularly interesting as it implies there is limited impact of the different soil state and impact of lateral boundary conditions in the later stages of the forecasts, or at least they do not dominate as strongly as potentially expected. This is investigated further in Section 4.

Location of convective events

Figure 10a shows the evolution of average number of points reaching the convective threshold. It indicates that WRF has the greatest activity and FV3 the least. The differences here are greater between regional models than driving models. The different diurnal cycles in the models are apparent and as such there will be a small influence on the interpretation of F_{common}. Figure 10b shows the evolution of the average fraction of common points from the start of the forecasts. Larger values imply greater agreement in location and smaller values imply more disagreement in location of convective points. As expected from the start of the forecasts F_{common} decreases with lead time when comparing the regional models as the forecasts diverge from each other, this divergence is due to a combination of different positioning of convective points but also the different number of convective points. On the other hand, spin-up factors are still detectable in the driving model comparisons as there is an initial increase in values (for up to 10 h) and then F_{common} starts to decay and level off as the influence becomes more confined to the lateral boundary conditions as opposed to the initial conditions.

© Crown copyright 2021, Met Office

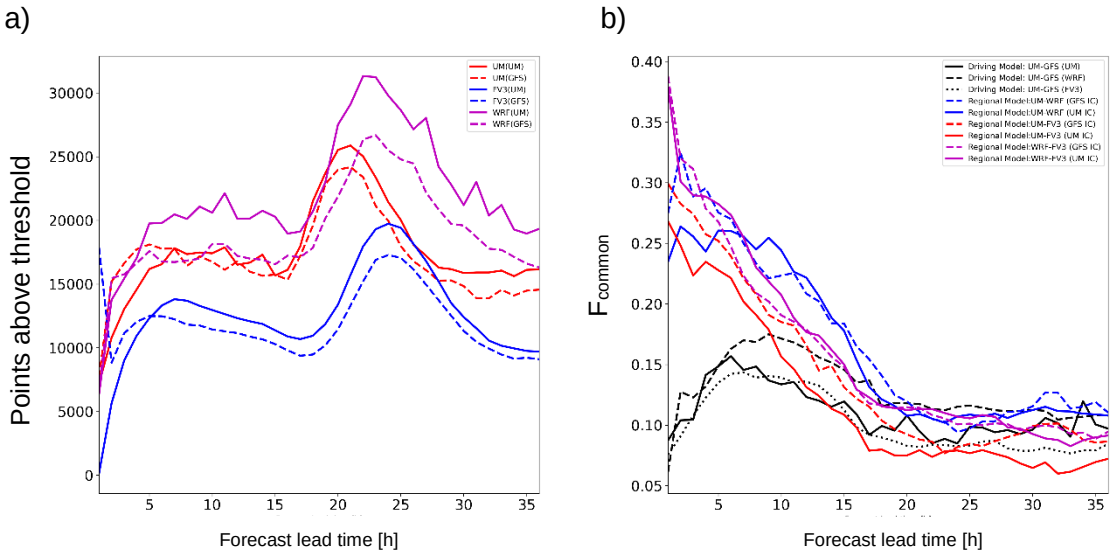


Figure 10: a) The average HWT number of points exceeding the convective threshold in all forecasts – the black line here represents sensitivity experiments which show very small differences between them, b) the average HWT fraction of common points throughout the forecast. Black lines represent driving model differences, and coloured lines represented regional model differences for UM-WRF (blue), UM-FV3 (red) and WRF-FV3 (magenta). A threshold of 30 dBZ is used to determine whether a point is convective.

In comparing the evolution and values of the different comparisons in Fig. 10b it is suggestive that the driving model dominates in the change in position early on in the forecast (lasting 10—12 h) before the regional model influences, via the evolution and development of convection, start to become just as important. This would be in line with the expected result (when thinking about similar results in convective-scale ensembles (e.g. Keil et al. 2014)). However, the dominance of the driving model occurs entirely during the initial shock phase where we are less confident of contaminating factors having influence. Therefore, whilst promising the results may not be robust. Figure 10b also suggests that the regional models have an influence in determining the frequency of precipitation and so cannot be completely ruled out from some of these differences either, although the number of convective points are more similar for WRF and UM compared to FV3.

Convective structure: CFI

Figure 11 shows the dependency on threshold of the CFI for the models and the observations and its evolution in time through a snapshot of events. Initially (Fig. 11a), and

© Crown copyright 2021, Met Office

most likely due to lack of spin-up of convective-scale features in the model, all models have a CFI that is larger than the observations suggesting the events are too large and there are not enough small scale features present. This behaviour is expected due to all the regional models being initiated from a ‘cold start’ or global model analysis. The global models do not have fine scale convective structure in them, reflected in the CFI scores for T+0 which would imply larger ‘clumps’ of convection in the models. A notable exception is at the higher dBZ thresholds, where the models show more scattered fragmented structures than the observations. This feature is true throughout the forecast and may be due to reduced frequency of observed cases above 60 dBZ, or potentially could indicate some other model bias.

After spin-up and initial shock are over most of the models increase their fragmentation and become closer to the observations (Fig. 11b). This behaviour is dampened in FV3 where the convection remains less scattered and more organised than the radar observations.

At T+18 to T+22, the UM regional model begins to overshoot the observations with convection at lower thresholds (30-40 dBZ) becoming more fragmented than the radar observations (Fig. 11c). This finding is of interest because previous studies on the subjective preferences of Operational Meteorologists have suggested that they prefer WRF to UM for the HWT domain. One hypothesis might be that these less organised light showers in the UM are particularly distracting for the human eye which is more likely to pick up the differences between clear sky (white background) and rain (blue objects) than the more subtle differences in magnitude of rainfall within the convection object. Hence, there should be some importance placed on models correctly simulating the ‘character’ of convection, somewhat quantified by this CFI score, as well as the amount and placement of convection (e.g. FSS and other skill scores).

Interestingly at T+30 the CFI in the UM returns to values closer to the observations (Fig. 11d). This coincides with the time where we would expect diurnal heating to be reduced and convection to decay.

A further notable factor in Fig. 11 is the evolution with increasing threshold. In the observations there is a curve towards increasing values of CFI and it suggests that the events become naturally organised at around 60 dBZ when the curve begins to plateau close to one. A similar behaviour, after spin-up, is apparent in the models (Figs. 11b and c). This natural organisation is not as strong, and at larger reflectivity values, suggesting

© Crown copyright 2021, Met Office

that there is a bias in the regional models. This can be confirmed to be the regional models over the driving models by the similarity of the models with the two driving conditions.

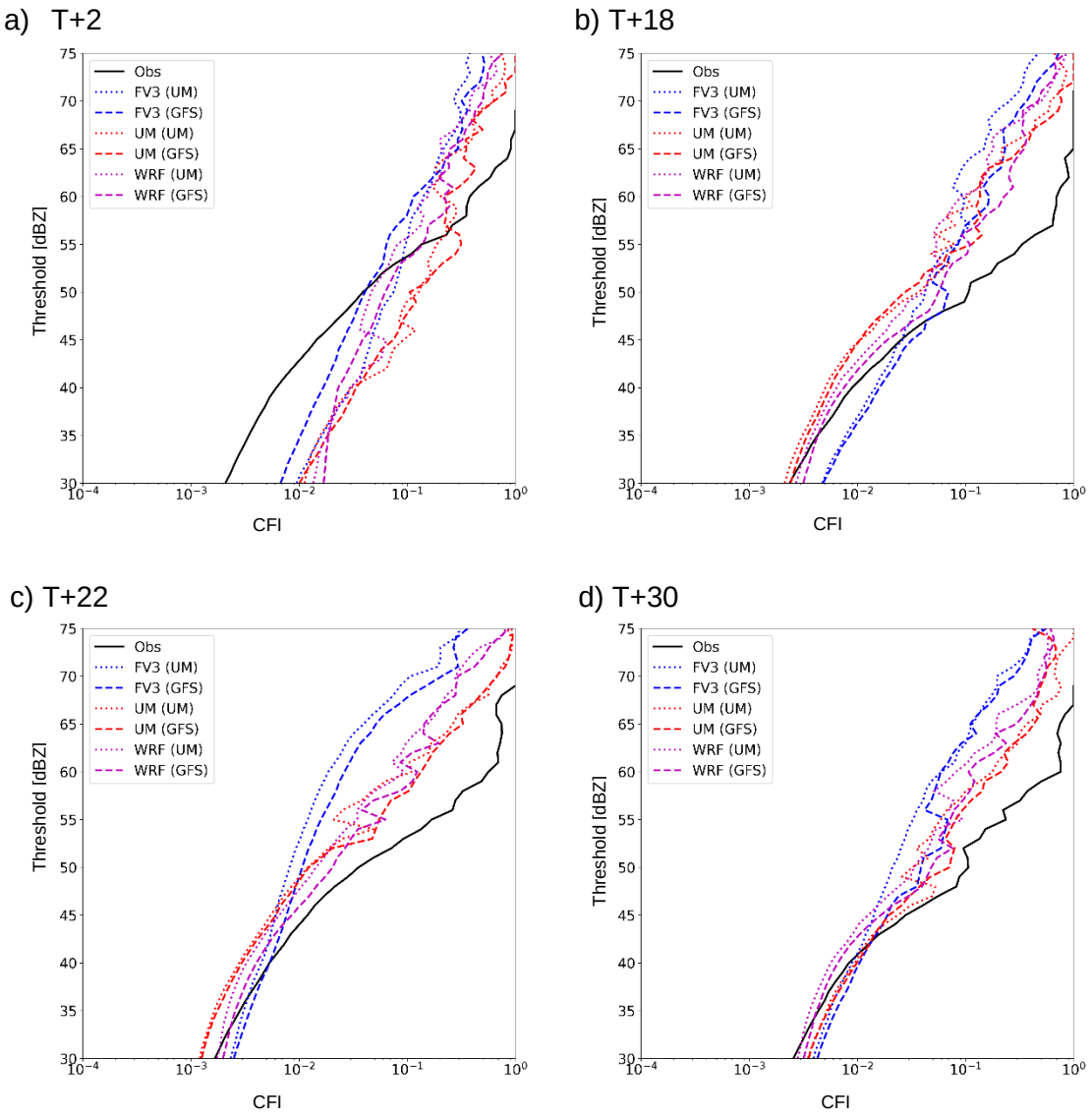


Figure 11: The average convection fragmentation index as a function of reflectivity threshold across the HWT for observations (black), FV3 (blue), UM (red) and WRF (magenta), with UM driving conditions (dotted) and GFS driving conditions (dashed) at a) T+2, b) T+18, c) T+22 and d) T+30.

Viewing the evolution of CFI with thresholds suggests it might also be useful to examine the evolution of the CFI at a single threshold. For the purposes of Fig. 12 a threshold of 30 dBZ was considered. A forecast lead time of T+16 onwards is shown to reduce the impact of shock and spin-up. The evolution confirms that there is a greater dependence on

© Crown copyright 2021, Met Office

regional model than driving model, and this is particularly true between T+18 and T+23 where there is a significant difference between the models at the 95% confidence interval.

The models (and observations) all show a reduction in the CFI as the convection begins to form, and the earlier diurnal cycle in the UM is shown by the earlier minima in the CFI compared to the other models. The reduction in CFI in the UM is stronger than in observations and the other models suggesting that the UM tends to initiate convection that is small (and thus strongly fragmented). Furthermore, the increase in the CFI is slower in the UM simulations compared to the other models suggesting that upscale growth of convection is relatively poorly represented in the UM, in agreement with the UM regional model configurations in other parts of the world (e.g. Keat et al. 2019). The FV3 forecasts tend to show too large systems in comparison to the observations (suggesting not enough breakdown) and the WRF is much more similar to the observations, but still indicating there is not enough fragmentation, particularly when considering forecast day two.

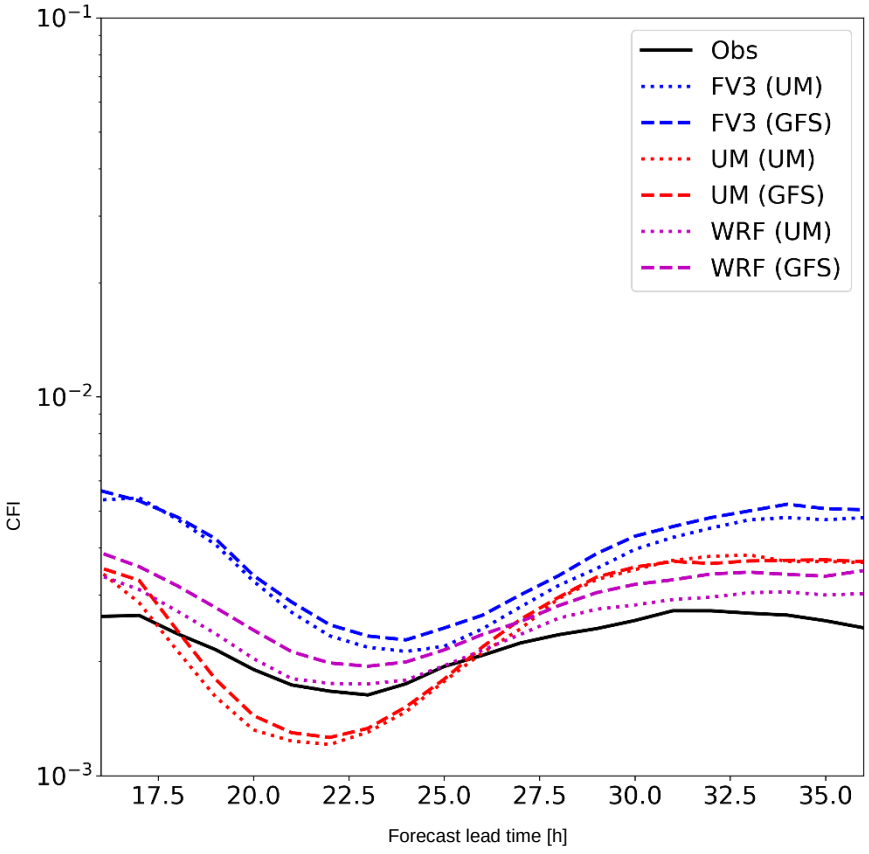


Figure 12: The average convection fragmentation index across the HWT as a function lead time after spin-up for a threshold of 30 dBZ. Observations (black), FV3 (blue), UM (red) and WRF (magenta), with UM driving conditions (dotted) and GFS driving conditions (dashed).

© Crown copyright 2021, Met Office

These results help quantitatively confirm the subjective analysis. However, with all these results caution must be applied given the soil state could be causing hidden differences (e.g. potentially on convective initiation and development).

At the peak of model convection initiation in figure 12 (illustrated by a dip to lower values of CFI in each line indicating scattered showers) there is some indication of larger differences between the driving model conditions for the UM and WRF, with WRF showing more impact from the driving model. Interestingly, FV3 shows this difference more when convection is more mature which would suggest the driving model is making a bigger influence on convection once it is at a more organised stage. However, these differences are not statistically significant. Furthermore, WRF and FV3 are using non-native soil state whilst the UM is using a consistent soil state for these runs and this may be an example where the soil is making an impact on convection initiation and development, although it is impossible to unpick with the information here.

Convective structure: Proportion of stratiform and convective precipitation

A further of aspect of the convective structure that can be considered, and was highlighted during the HWT SFE discussions, is the reduction in stratiform rain in the UM compared to the other models. Figure 13 shows 2D histograms of the area of the convective cells against the CP. The observations show a strong tendency for relatively small weak cells but also the larger cells have a lower CP. There are occasional small intense cells in the observed values as well. In comparing the different columns in Fig. 13 by eye there is a larger difference down the columns (between regional models) than across the rows (between driving models).

Both UM forecasts appear to have a structure that is closest to the observations, although there are, perhaps, more small intense storms than the observations suggesting a lack of stratiform regions in the storms. The FV3 forecasts show a stronger tendency for larger storms with large stratiform components; the WRF forecasts have a strong tendency for larger storms with a large stratiform component.

The lack of differences between the driving models is clearer in Fig. 14 and suggests that on average the UM driving conditions may lead to more smaller storms with a small convective component than the GFS driving conditions. However, as previously discussed

© Crown copyright 2021, Met Office

and noticed during subjective assessments the larger differences are in the regional models.

Figure 15 shows the regional model differences. The UM tends to have more small storms with large CPs than the other two models but fewer large storms with small CPs. Further to this FV3 has more small convective events with larger CPs than WRF. WRF has more small-medium convective events with CPs ranging between 0 and 0.2.

As with the CFI these results closely match the subjective views during the HWT. Furthermore, despite the differences in soil state between the runs there does not, immediately, appear to be an influence on the convection. This difference could be more subtle, as potentially in the CFI, and the soil may have an impact during the initiation phase of convection.

© Crown copyright 2021, Met Office

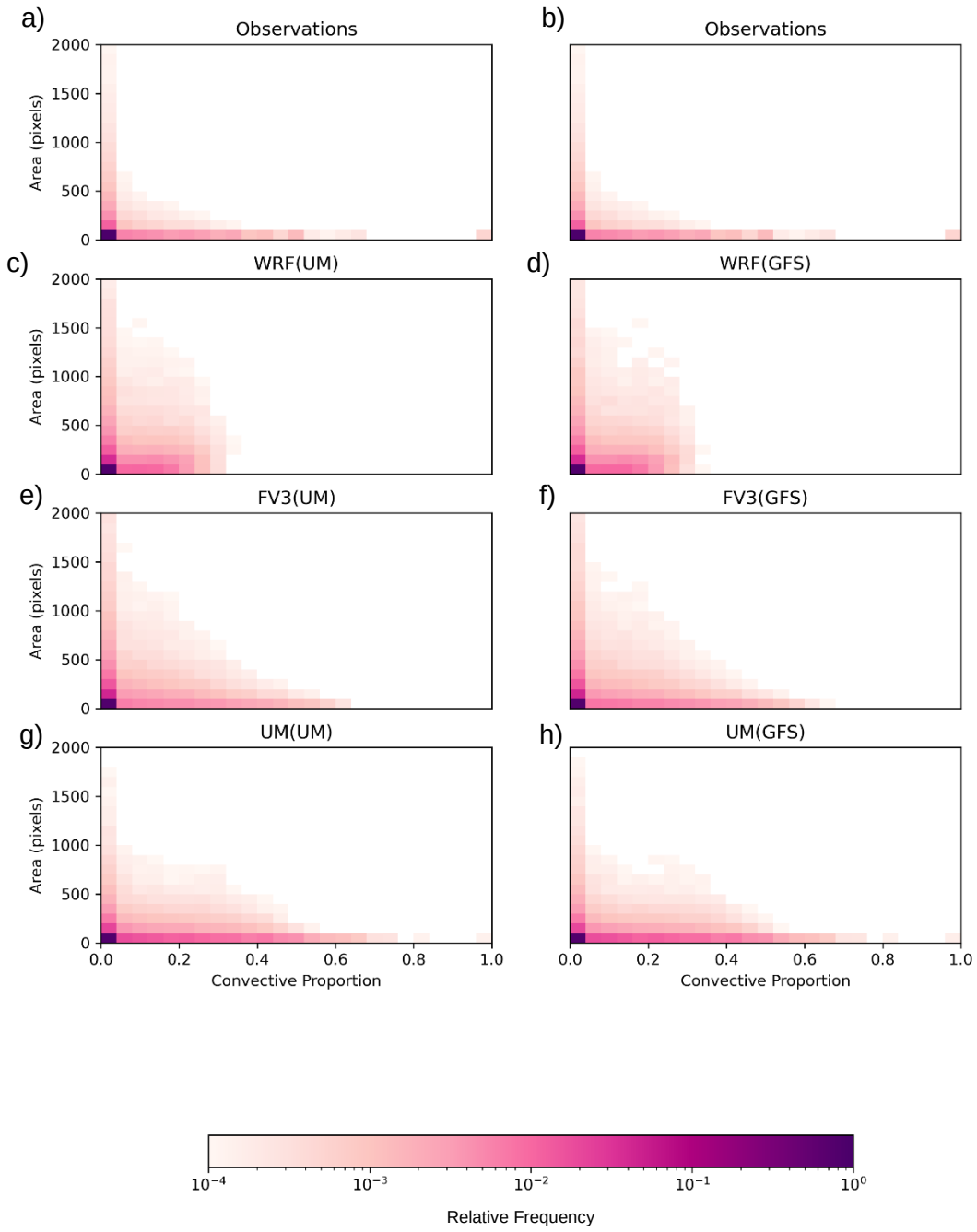


Figure 13: 2D histograms of convective proportion *vs*. area of convective objects for a and b) Observations, c) WRF(UM), d) WRF(GFS), e) FV3(UM), f) FV3(GFS), g) UM(UM) and h) UM(GFS). All data after spin-up is considered across the entire HWT.

© Crown copyright 2021, Met Office

Summary of objective results

Initial shock combined with spin-up lasts on average 15 h and is clearly detected in the temperature variance plots. These show two factors influencing the early results of the experiments: the shock in driving model comparisons and the soil state in regional model comparisons (see Section 4.1 for more details). These differences indicate a need to consider results beyond the first 16h lead time in events for assessing the relative importance of the driving model and regional model.

Histograms, after spin-up and shock, show no statistically significant differences between forecasts for any of the variables considered, although different windspeed and reflectivity characteristics are apparent between different regional models. The lack of significant differences suggests that it is plausible to create an ensemble out of these simulations. Results on location and convective structure appear to be more promising with location being dominated by the driving model early on (but this is during the shock phase) and after the shock phase it is unclear. The convective structure, which is examined only after spin-up and shock, indicates that the regional model has greater influence on the structure of the convection and is likely a result of the different microphysics schemes and boundary layer structures.

The sensitivity of these results to soil state and LBCs are examined further in the next section.

© Crown copyright 2021, Met Office

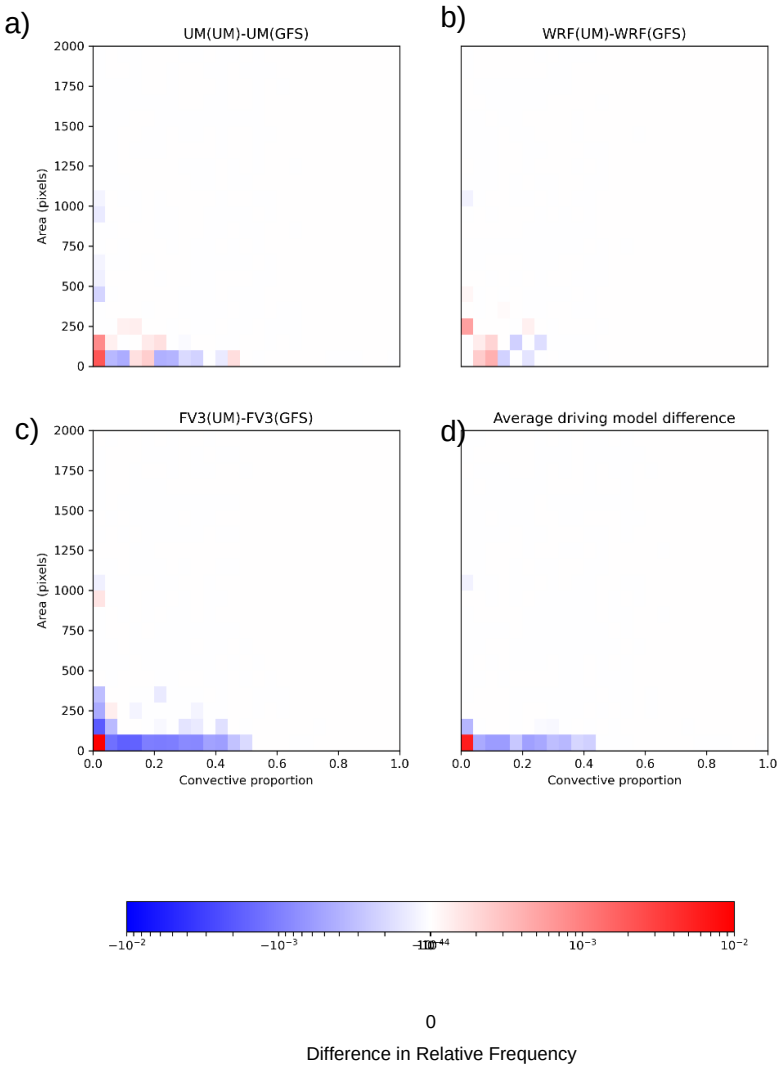


Figure 14: Differences in the 2D histograms presented in Fig. 13 for the driving conditions, a) UM(UM)-UM(GFS), b) WRF(UM)-WRF(GFS), c) FV3(UM)-FV3(GFS) and d) the average of a, b and c. Reds indicate the UM driving conditions populate this area more and blues indicate the GFS driving conditions populate this area more.

© Crown copyright 2021, Met Office

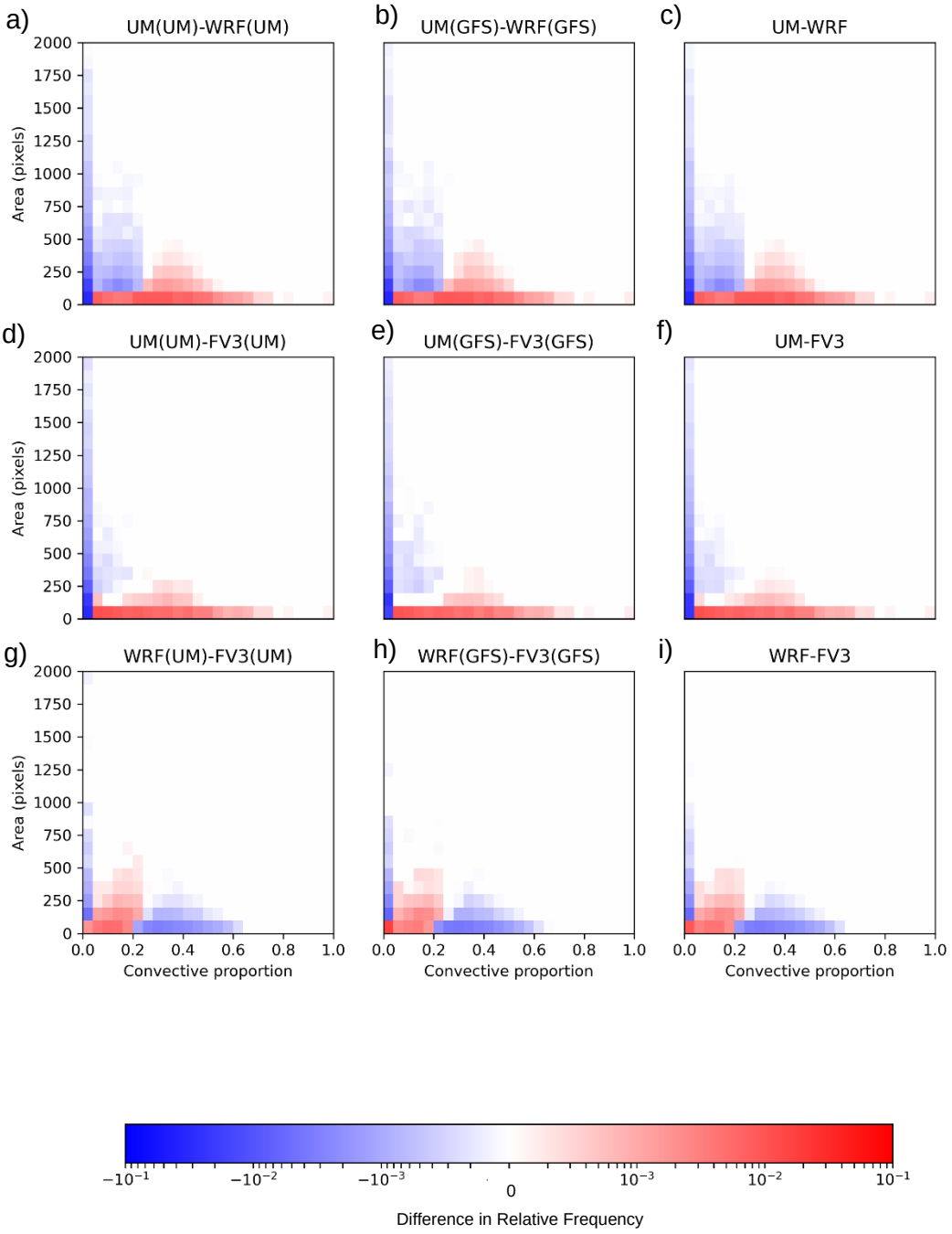


Figure 15: Differences in the 2D histograms presented in Fig. 13 between regional models, a,d,e) are differences with UM driving conditions, b,e,h) differences with GFS driving conditions and c,f,i) are the average between the two; a,b,c) UM-WRF differences, d,e,f) UM-FV3 differences and g,h,i) WRF-FV3 differences. Reds represent areas where the first model populates the histograms more and blue areas where the second histogram populates the distribution more.

© Crown copyright 2021, Met Office

4. Sensitivity to experimental set-up

Sensitivity to model set up was examined through two control tests using the WRF simulations and one accidental test using the UM regional model. The WRF sensitivity tests occurred by altering the WRF(UM) simulations such that one simulation used GFS lateral boundary conditions instead of UM lateral boundary conditions, and the other experiment used GFS soil state instead of UM soil state. The latter is equivalent to the methodology in place in the TRANSPOSE-AMIP experiments (Williams et al. 2013). The results from these sensitivity tests, using all the previous diagnostics are presented in Figs. 16—21. The impact of the soil state is discussed in Section 4.1; the impact of the lateral boundary conditions is presented in Section 4.2. The accidental experiment with GFS soil temperature in the UM is described in section 4.3 and further considerations for these experiments are discussed in Section 4.3.

4.1. Soil state

In previous literature on these types of experiments, especially for climate timescale experiments, the soil state (at the very least soil moisture) is kept consistent between the (regional) model runs with different driving conditions. There are two reasons behind this: i) soil moisture is not well constrained as there are limited observations and this lack of constrain results in poor analyses of soil moisture (e.g. Keil et al. 2019); and ii) soil moisture is not consistently defined in models (so model output of soil moisture cannot reliably be compared between models). There is scope to change the soil state with changing initial conditions, but this comes with a requirement of not using the data directly from the driving model's analysis. Instead the analysed soil states are run in the (regional) model's land surface model for several months so that it can adjust to the same model definitions (e.g. Phillips et al. 2004) or nudged in using Boyle et al.'s (2005) method, of which the former is the preferred method.

© Crown copyright 2021, Met Office

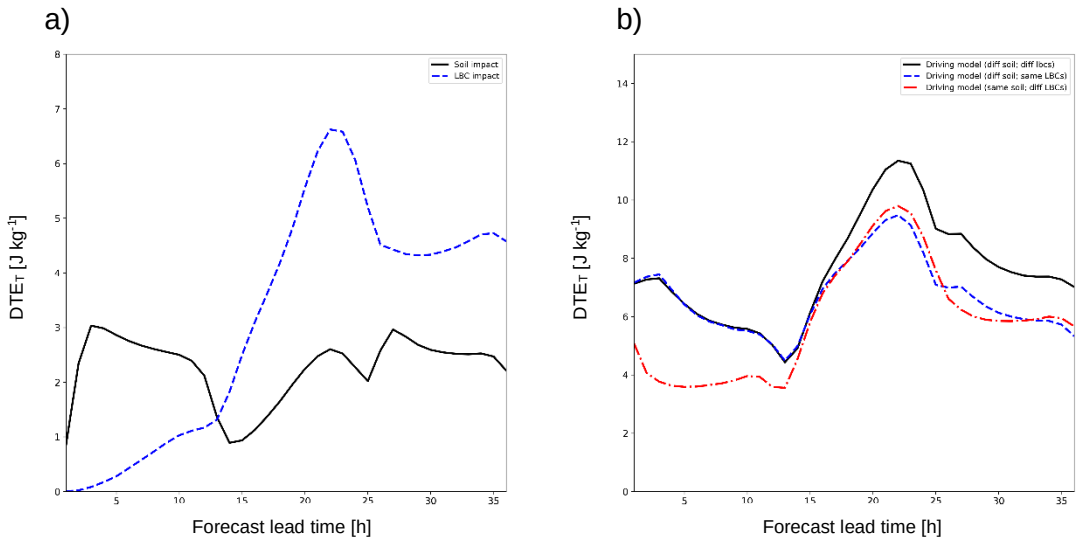


Figure 16: Average temperature variance evolution across the HWT for a) WRF(UM)-WRF sensitivity experiments (blue change in lateral boundary conditions and black change in soil state) and b) comparisons against the driving model simulations with the different sensitivity tests: WRF(UM)-WRF(GFS) (black), WRF(UM; GFS LBCs)-WRF(GFS) (blue) and WRF(UM; GFS SOIL)-WRF(GFS) (red).

Figure 16a shows the behaviour when only the soil state is changed between a model run. There is a visible diurnal cycle, and the evolution that is seen in the temperature variance is replicated in the model comparisons of the temperature variance in Fig. 7b (most notably in the WRF(UM) and FV3(UM) simulation comparisons). This similarly between figure 16a and figure 7 implies that the DTE differences between regional models may be strongly influenced by the soil moisture. This makes intercomparisons between the models challenging.

The difference is particularly notable at the start of the run, compare Fig. 16b black line (driving model differences with different soil temperature) and red line (driving model differences with the same soil temperature). This large difference indicates that part of the initial shock being detected in the models is due to soil moisture. There is still some difference after most of the shock has subsided. During the shock phase the difference is as much as a 1 K change in surface temperature, whereas it reduces to approximately 0.3 K the following day.

Particularly during the shock phase, the question is whether this impact could change the convective temperature and stability of the atmosphere as well as the intensity of the convection. The impact will be somewhat reduced for the second day of the forecast when the impact is one third of the initial shock. It is worth noting that in the analysis that considers points after the shock there appears (on first glance) to be less of an impact (Figs. 17—21).

© Crown copyright 2021, Met Office

However, there is a lasting difference in Fig. 16b and it appears that both the LBCs and the soil moisture are partially influencing results to a similar degree in the latter parts of the forecast. As such there is evidence that soil moisture should be considered carefully in these experiments and the models show sensitivity to soil state throughout the forecast, in agreement with Keil et al. (2019).

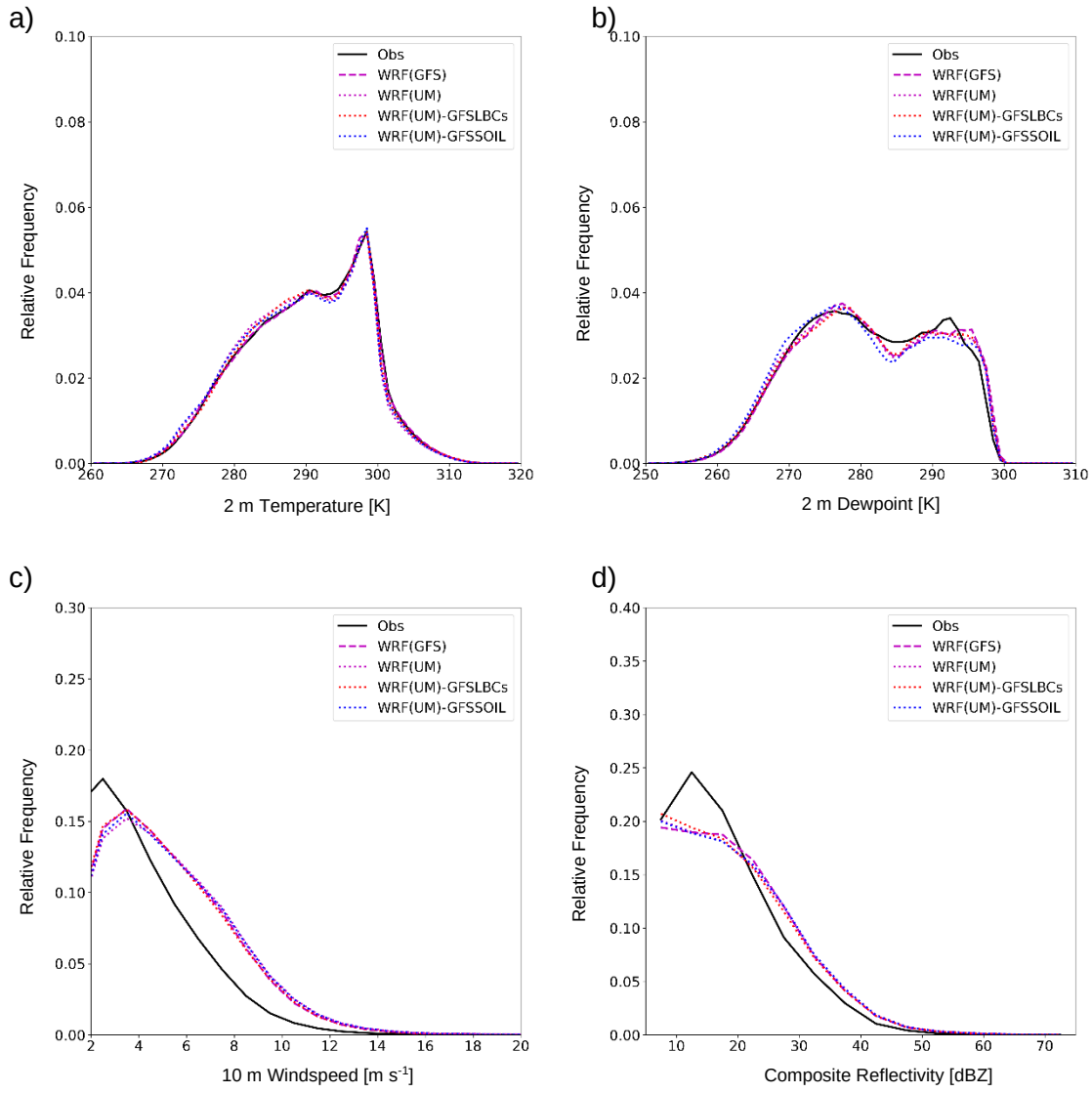


Figure 17: Histograms for surface variables during the HWT for the WRF sensitivity experiments, observations (black), WRF(GFS) (magenta dashed), WRF(UM) (magenta dotted), WRF(UM: GFS LBCs) (red dotted) and WRF(UM: GFS SOIL) (blue dotted) for a) 2 m temperature, b) 2 m dewpoint temperature, c) 10 m windspeed and d) composite reflectivity.

© Crown copyright 2021, Met Office

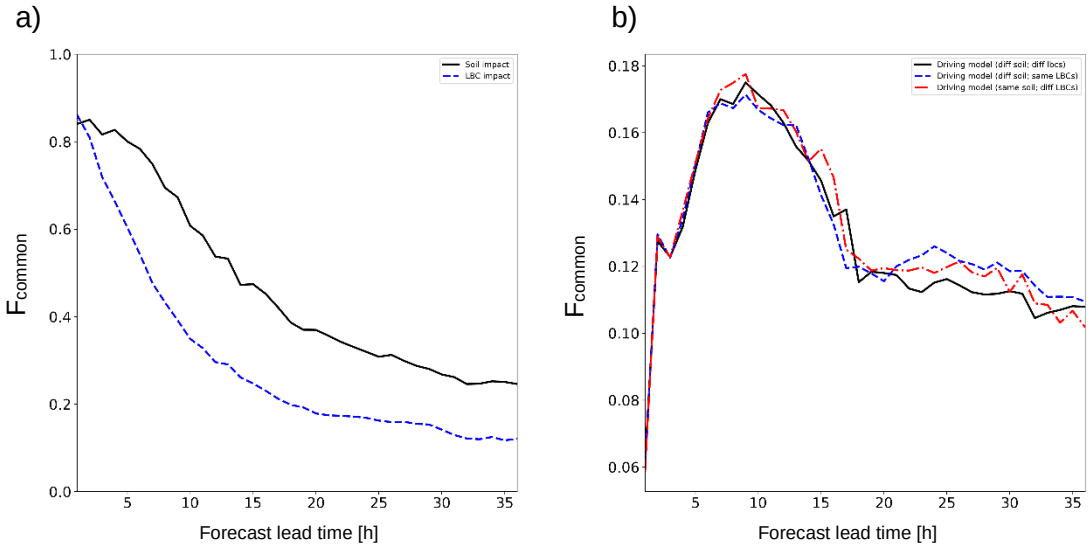


Figure 18: Average fraction of common points across the HWT for a) WRF(UM)-WRF sensitivity experiments (blue change in lateral boundary conditions and black change in soil state) and b) comparisons against the driving model simulations with the different sensitivity tests: WRF(UM)-WRF(GFS) (black), WRF(UM; GFS LBCs)-WRF(GFS) (blue) and WRF(UM; GFS SOIL)-WRF(GFS) (red). The number of points reaching the convective threshold (not shown) remains in between the two WRF curves (and are closer to the WRF(UM) curve) shown in Fig. 10a.

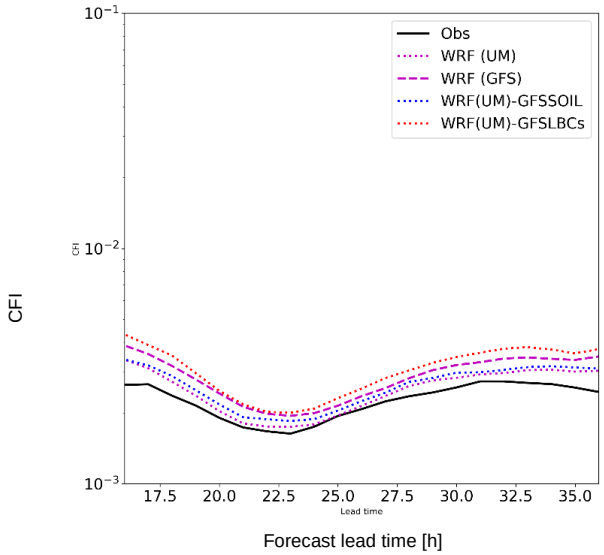


Figure 19: The average convective fragmentation index during the HWT for the WRF sensitivity experiments, observations (black), WRF(GFS) (magenta dashed), WRF(UM) (magenta dotted), WRF(UM: GFS LBCs) (red dotted) and WRF(UM: GFS SOIL) (blue dotted). Only data after spin-up is presented.

© Crown copyright 2021, Met Office

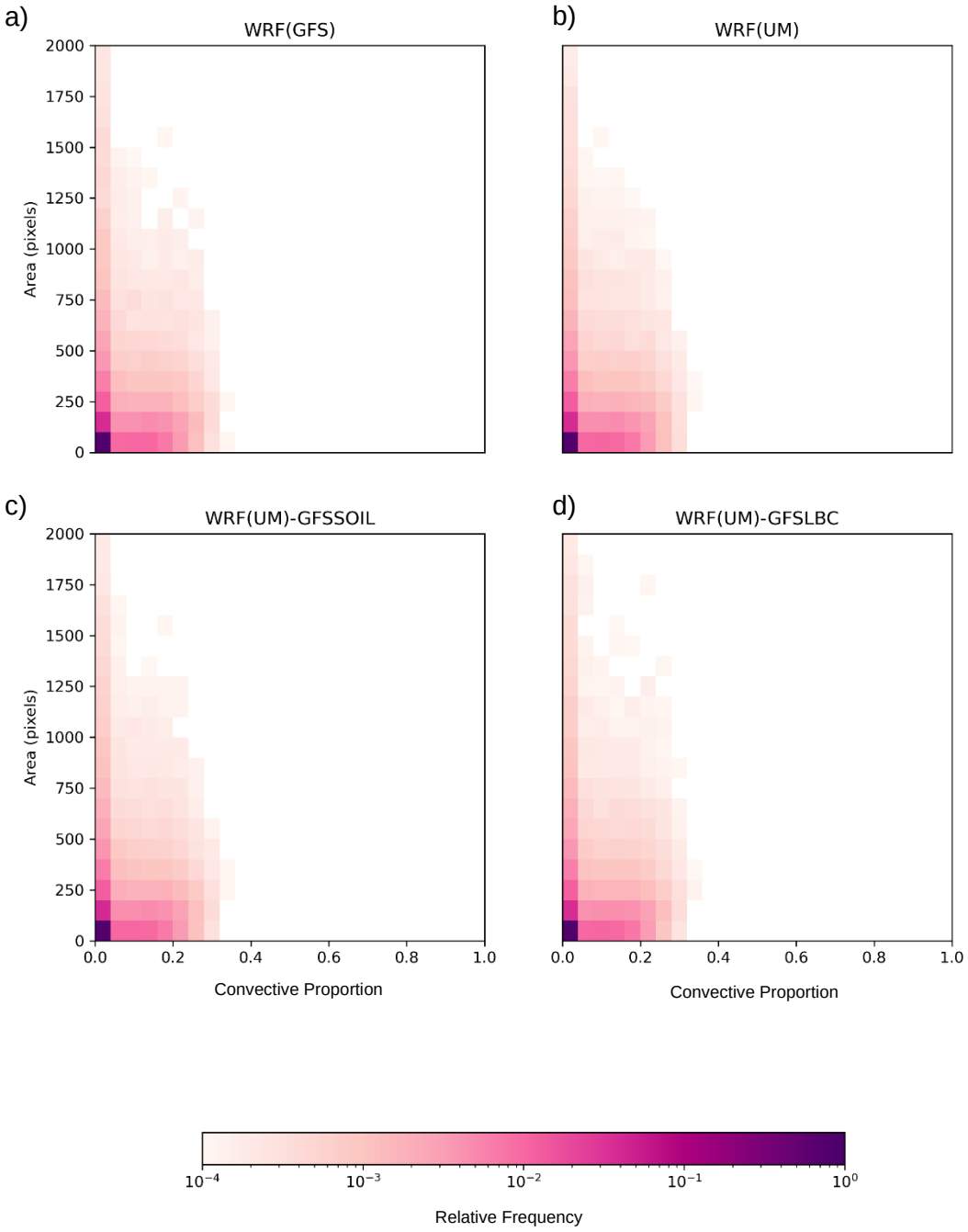


Figure 20: 2D histograms for the convective proportion *vs*. area of convective events for the entire HWT (after spin-up) for the WRF sensitivity experiments a) WRF(GFS), b) WRF(UM), c) WRF(UM: GFS SOIL) and d) WRF(UM: GFS LBCs).

© Crown copyright 2021, Met Office

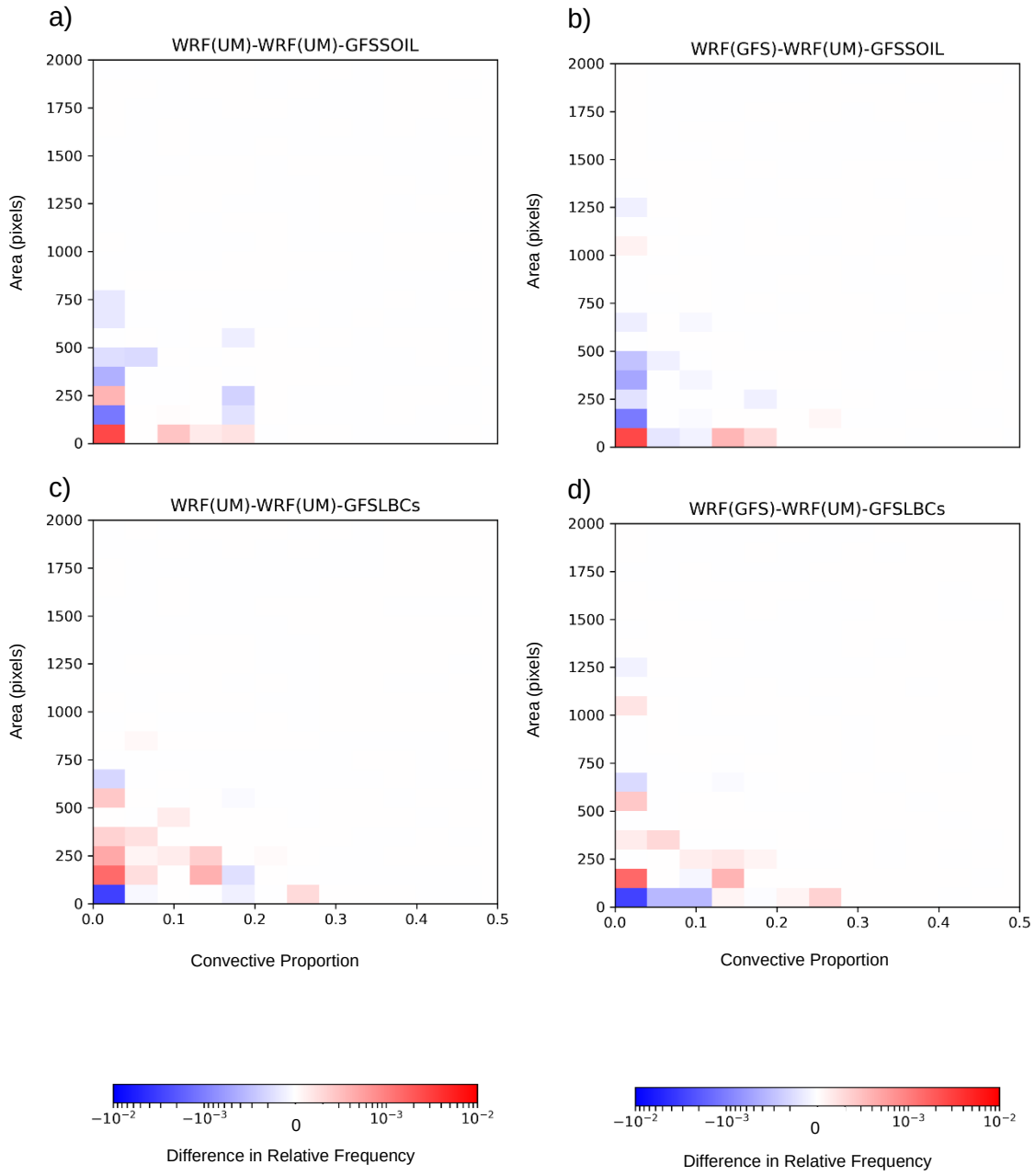


Figure 21: Differences between the 2D histograms presented in Fig. 20, a) WRF(UM)-WRF(UM: GFS SOIL), b) WRF(GFS)-WRF(UM: GFS SOIL), c) WRF(UM)-WRF(UM: GFS LBC) and d), WRF(GFS)-WRF(UM: GFS LBC). Panels b and d should be compared against Fig. 14b. Red colours imply that the first model populates this area more and blue colours imply the second model populates this area more.

The subtle impacts on the convection itself, whilst not appearing obvious, are important differences that can be seen within the model and across the model comparisons. It is worth noting that they do not appear to qualitatively change the impact of what matters more in terms of driving *vs.* regional model. However, there are subtle impacts on storm locations, structure/fragmentation and size/intensity. Given that differences between all

© Crown copyright 2021, Met Office

experiments are small, these subtle differences do cause an issue for the analysis and cannot be disregarded as it becomes difficult to disentangle the influencing factors on convection.

The number of convective points does not change between the changes of soil state (not shown). This means any differences in F_{common} result purely from a change in location of convective events. Figure 18a shows that by the end of the spin-up and shock period there has been a substantial movement of convective cells (less than 40% of points are in the same location compared to the same run where the change is only occurring in soil state). There is a reduced impact in comparing with the forecasts based on different driving conditions (Fig 18b) but this impact shows that there is greater agreement between the positioning of the two forecasts after initial shock.

There is also a reduction, that is equivalent to changing the LBCs, in the fragmentation of the convective events (bringing them more in line with observations; Fig. 19) thus the convective structure is changed. This is confirmed through the impact on the CP and area histograms (compare Figs. 14b and 21b; note the colours are reversed between the two figures) which show that with the native soil there is a greater reduction in the size and intensity of the convective precipitation than when non-native soil moisture is used.

Given the impact of the soil state shown here, and based on previous literature, soil state does influence the results quantitatively, if not qualitatively, and as such cannot be neglected in an objective analysis. Therefore, soil state is important for these simulations and needs to be consistent throughout the experiments with the same regional model to allow for robust conclusions to be drawn. The reduced qualitative difference is likely due to only areas outside of the main model shock (in which true model shock and the artificial inclusion of soil shock is combined) is considered.

4.2. Lateral boundary conditions

The lateral boundary conditions have also been considered to see if the impact of the initial conditions as opposed to just the driving model can be determined. As with the soil state changes, the greatest difference occurs in the temperature variance (Fig. 16). The comparisons showing just the impact of changing lateral boundary conditions (Fig. 16a) show expected behaviour of there being a greater influence from the lateral boundary conditions in the leadup to day two of the forecast. This is also reflected in Fig. 16b with the results from day one being identical to that of the experiment used for comparisons in Section 3. This implies that the initial conditions have an impact lasting to approximately 16

© Crown copyright 2021, Met Office

hours (when the results begin to diverge (though this is masked by the initial shock in WRF which lasts for 12—15 hours). It is worth noting that towards the end of day one and into day two the lateral boundary conditions appear to have just as much impact as the soil state.

Qualitatively the impact of the lateral boundary conditions is small in Figs. 17—21. This lack of difference could feasibly be due to the analysis domain being away from the boundaries, and outside the range of boundary spin-up, for this model.

The results from this sensitivity experiment looking at lateral boundary conditions indicate that the overall results of this experiment are less sensitive to the change in lateral boundary conditions compared to changes in the soil moisture. However, there are still sensitivities exhibited. These sensitivities suggest that the main idea behind this experiment is better phrased in terms of driving model *vs*. regional model (as we have done throughout the report) as opposed to purely initial conditions *vs*. model core (as termed during the HWT SFE).

4.3. Accidental experiments with soil temperature

The role of soil temperature, whilst being more constrained, is also important. A mistake was made with an early configuration of the UM regional model driven by GFS in that all the soil temperature layers were set to the 2m soil temperature. This obviously gives an unrealistic soil temperature at the surface and the mid layers. This mistake was corrected to the appropriate GFS soil temperature values at the correct levels. However, it also allows the comparison of the two runs to inspect the impact of initial condition soil temperature on temperature and humidity at the surface and in the soil.

Figure 22 shows the impact on some key model variables throughout the forecast. Surprisingly for such a major shock to the temperature, there is not too much difference in the 1.5m temperatures in the latter stages of the forecast. There is more impact on the 1.5m dewpoint, with the influence proportionate to the diurnal cycle. There is more direct influence on the soil temperature and moisture, especially at level 2 which is less moderated by the atmosphere. Figure 23 shows the experiments compared to the other regional model runs and large differences can be seen, demonstrating an extreme reaction in DTE. This will likely, although not shown here, have impacts on the structure and likelihood of convection based on the differences between reaching the convective temperature.

© Crown copyright 2021, Met Office

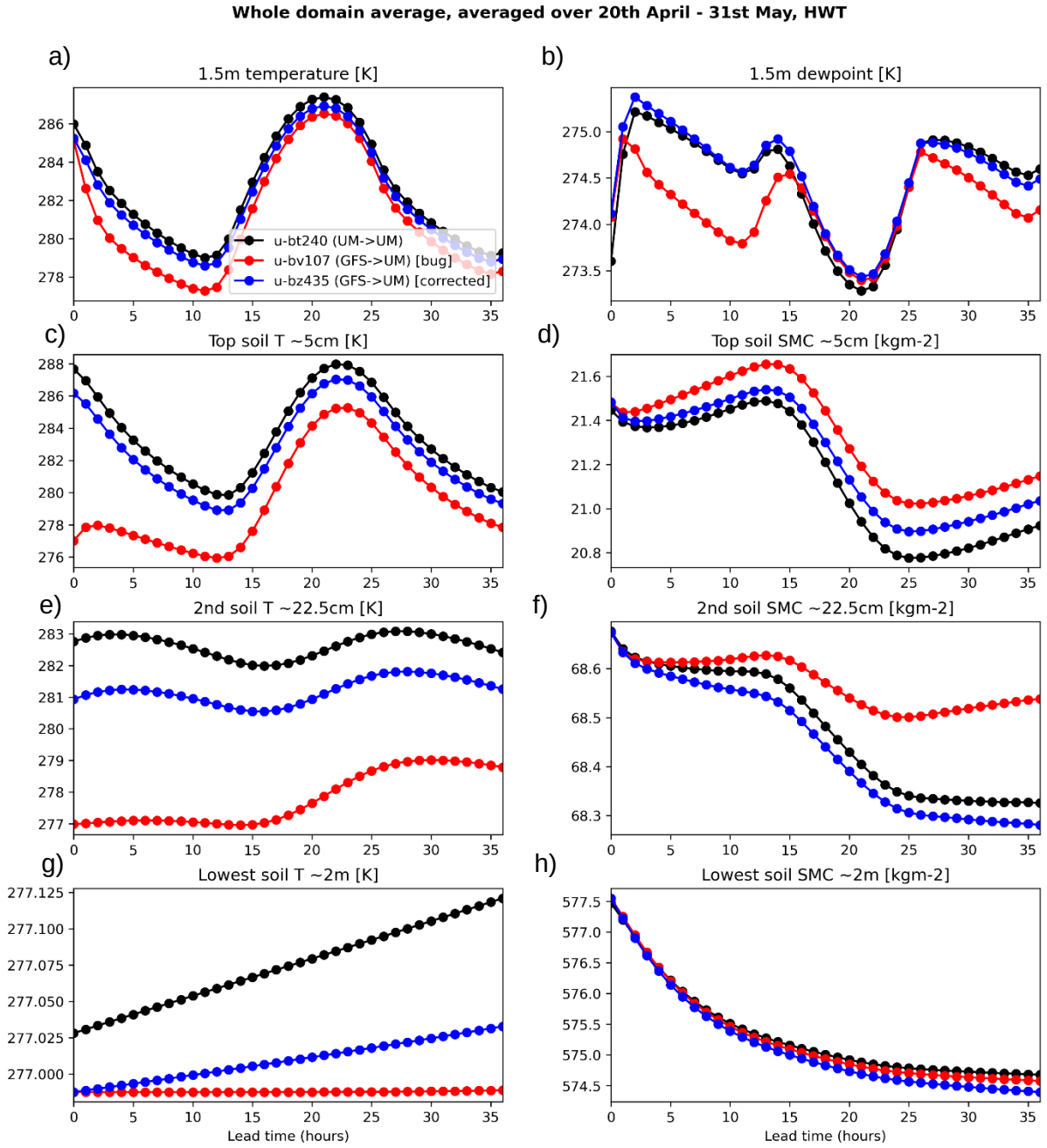


Figure 22: Soil differences within the different UM simulations. Black lines show the native soil temperature and moisture, red lines show the UM(GFS) with incorrect soil temperature interpolation and the blue lines show the UM(GFS) with corrected soil temperature interpolation. The impact on the surface temperature and dewpoint as well as the soil temperature and soil moisture at different depths throughout the forecasts as an average across the entire HWT period.

© Crown copyright 2021, Met Office

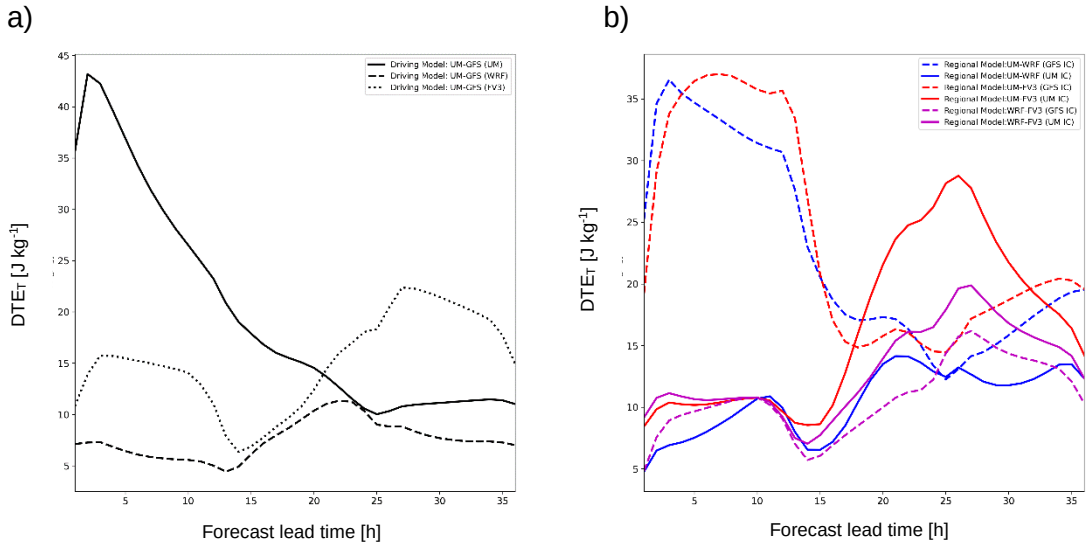


Figure 23: As for Fig. 7 but with soil temperature fixed at 2 m depth throughout the soil profile in the UM(GFS) simulations.

4.4. Other considerations

There are other factors that need to be considered that are not addressed in this current experiment and these are summarised here:

- Inconsistency of runs between the different centres – ideally the same domains, consistent methodology of soil moisture states should be applied
- Impact of initialisation (due to lack of all models producing T+0 data).
- Impact of different specification of vertical resolution of driving data.
- Impact of domain size (due to different boundary spin-up).
- Impact of different (effective) resolutions (although this will be small as all are convection-permitting models with grid lengths of the same order of magnitude).
- Forecast length may not be long enough, given the length of spin-up and shock, to allow model differences to be truly, and fairly, detected.

Given these caveats, several aspects of this experiment can be improved upon for future versions of these experiments, and the lessons learnt and recommendations for such future experiments are described next.

© Crown copyright 2021, Met Office

5. Recommendations for the Future

In this section recommendations based on the results from HWT SFE 2020 are made that will lead to the improvement of this experiment, and produce clearer conclusions on the scientific questions around the evolution of the relative importance and determination of the cause or location of errors in the regional and driving models. The four main recommendations on soil state (Section 5.1), domain size (Section 5.2), driving conditions (Section 5.3) and initialisation time (Section 5.4) are discussed in detail together with some additional recommendations (Section 5.5) briefly discussed.

5.1. Soil state

The results from Section 4, and in comparison with those in Section 3 have indicated that the soil state has an influence on convection and should not be neglected for a convection-based study, in agreement with Keil et al. (2019). One of the main reasons behind the soil state being kept consistent in historical experiments of this type is that soil moisture is not well constrained and is often defined differently in different models. In the TRANSPOSE-AMIP protocol (e.g. Phillips et al. 2004, Williams et al. 2013) it states that a soil state should not be used directly from the analysis of different models (i.e. soil states should be consistent between the same regional model). Therefore, the preferred method is to use the same soil state for all simulations in each regional model so that WRF, FV3 and UM maintain their own soil state throughout.

5.2. Domain size

A consistent domain size should ideally be used for this type of experiment. This is to ensure that if sub-domains are used they are all at the same distance from the lateral boundaries. Currently the domains are of different sizes (Fig. 1) and as such a common domain of the entire UM domain has been used to ensure fair comparisons across the models. However, this choice of comparison domain leads to the UM forecasts being open to boundary effects which WRF and FV3 are more sheltered from. Had a buffer been applied to the UM domain to reduce UM boundary impacts there would still be a problem based on different distances from the boundaries and different lengths of boundary spin-up. A consistent domain would reduce these issues.

© Crown copyright 2021, Met Office

5.3. Driving conditions

In a more ideal experiment, both driving models would specify output on the same vertical levels. This would ensure that no extra detail is given to one model, as opposed to the other, removing any unfair advantage and allowing the correct detection of where any errors lie, either in the regional model or in the driving model, as the result could be an impact of differently specified initial conditions.

The difference in driving conditions is known to be important (e.g. Porson et al. 2019). However, the vertical specification of the initial conditions is likely to show variability depending on the phenomenon examined, therefore it is safer to use the same vertical (and ideally horizontal) resolution of driving conditions for all simulations.

5.4. Initialisation time

The initialisation time should be included for all runs so that a true idea of differences in the initialisation state and its impact on the subsequent forecast can be detected. It also allows a greater idea of whether initial shock occurs in the forecasts, and how important it is to the forecast evolution.

5.5. Other recommendations

- A consistent methodology is set at all centres to allow for comparisons between all the model runs
- More live-run testing of models to ensure they are all available for subjective analysis during the HWT SFE.
- Ideally same horizontal resolution of regional models (Or at least horizontal resolutions are of the same order of magnitude and are converted onto the same grid for analysis)
- Same output levels are used for analysis
    - Similar output levels (1.5 vs. 2 m) are reasonable for surface parameters, but it would be better to have all on the same level. This factor becomes a requirement if looking at 3D fields (for which outputs interpolated onto pressure levels would be the best output to make the fairest comparisons).
- Increased forecast lengths
    - It would be good to consider longer forecasts lengths as, given the initial shock lasts for varying lengths of time in the different models, robust

© Crown copyright 2021, Met Office

analysis may only be able to be performed on the last few hours in which there is only limited data.

6. Summary

During the Hazardous Weather Testbed Spring Forecasting Experiment (HWT SFE) 2020 the Met Office contributed towards two experiments: an ensemble experiment and a deterministic experiment. The deterministic experiment is the focus of this report.

The deterministic experiment, originally, aimed at determining the impact of the driving model on the regional model for forecasts of severe convection and how the relative importance of each model evolves overtime. This experiment had a positive reaction and strong engagement during HWT SFE 2020. In the past these experiments have resulted in the ability to detect, and improve, model errors from parametrizations or specific regions (e.g. Williams et al. 2013), give indications on physical processes in models and how they different between models (e.g. Flack et al. 2021b) and also the impact on the spread of ensembles (e.g. Porson et al. 2019).

The initial analysis presented in this report have focussed on the question of the relative importance of driving model and regional model for severe convection. Useful results, from the latter part of the forecast (T+20 hrs), were obtained indicating that the convective structure (in particular, the fragmentation of convective cells and the ratio of convective to stratiform precipitation) appeared to be dominated by the regional model. They also indicated that the driving model had a larger impact on the position of convection in the early stages of the forecast but later had an equal weight with the evolution from the regional model. These results strongly agree with equivalent ideas in convective-scale ensembles where model physics perturbations dominating during convective initiation, initial condition perturbations at the beginning of the forecast and lateral boundary condition perturbations have a steady impact, particularly on the position of convection (e.g. Keil et al. 2014, Kühnlein et al. 2014, Flack et al. 2018).

This report has also clearly set out the caveats to the current iterations of these experiments. The main concern is the impact of using non-native soil moisture in the regional models. This was shown to create a noticeable shock, detectable into the second day of the forecast and also have some influence on convection location, structure and size/intensity.

Following this experiment, a set of four key recommendations were made for future experiments:

© Crown copyright 2021, Met Office

1. soil state should be consistent across the regional model, or at least soil moisture as it is not as well constrained as soil temperature;
2. the domain size should be identical ideally;
3. driving conditions should be specified on the same vertical levels to not introduce dependences based on vertical resolution;
4. initialisation time should be included in all simulations to give an idea of the impact of interpolation and a more complete idea of the initial shock.

The collaboration on this work is ongoing and it is hoped that we may create a more consistent set of experiments for further analysis (outlined in table 2). Results from these re-runs are then likely to be more intercomparable.

Therefore, whilst more work needs to be done in this area the potential use of this type of experiment has been shown and, given further improvement to the experimental protocols used in future iterations, , it has the potential to be a very powerful tool for model development and evaluation in the future. The evaluation metrics applied here should also provide benefit when applied more widely for similar model intercomparison exercises in future.

© Crown copyright 2021, Met Office

Table 2: Possible model setups for this experiment with comments. Underlined experiments are those that are used in this report. The ideal simulations (not included in the table) for the non-native driving model uses the regional model's native soil state.

Experiment ID	Regional Model	Driving Model	LBCs	Soil Moisture	Soil Temperature	Comment
WRF(UM)	WRF	UM	UM	UM	UM	Not optimal as soil moisture is not directly transferable between models
WRF(UM)	WRF	UM	UM	GFS	UM	Compromise – but robust as soil T is well constrained (preferred)
WRF(GFS)	WRF	GFS	GFS	GFS	GFS	Native setup (preferred)
FV3(UM)	FV3	UM	UM	UM	UM	Not optimal as soil moisture is not directly transferable between models
FV3(UM)	FV3	UM	UM	GFS	UM	Compromise – but robust as soil T is well constrained (preferred)
FV3(GFS)	FV3	GFS	GFS	GFS	GFS	Native setup (preferred)
UM(UM)	UM	UM	UM	UM	UM	Native setup (preferred)
UM(GFS)	UM	GFS	GFS	UM	GFS	Compromise – but robust as soil T is well constrained (preferred)
UM(GFS) – had during HWT	UM	GFS	GFS	GFS	GFS	Not optimal as soil moisture is not directly transferable between models

© Crown copyright 2021, Met Office

Acknowledgements

This work would not have been possible without the help of Stuart Webster, Paul Earnshaw, and Bjoern Fock for helping setup the UM(GFS) forecasts. We also wish to acknowledge all participants of HWT SFE 2020 for providing valuable comments and discussions during the HWT. Further thanks on the discussions at the Met Office go to Duncan Ackerley, Martin Best, Douglas Boyd, Mike Bush, Andrew Bushnell, Dan Copsey, Mike Cullen, Richard Keane, Humphrey Lean, Jon Petch, Aurore Porson, Nigel Roberts, Alistair Sellar, Martin Willet, David Walters, and Prince Xavier.

References

Bony, S., Bellon, G., Klocke, D., Sherwood, S., Fermepin, S., and Denvil, S. (2013) Robust direct effect of carbon dioxide on tropical circulation and regional precipitation, *Nature Geosci.*, **6**, 447–451, https://doi.org/10.1038/ngeo1799.

Bowen, G. W. and Burgess, R. L. (1981) A quantitative analysis of forest island pattern in selected Ohio landscapes. ORNL Environmental Sciences Division, Publication No. 1719, ORNL/TM7759. Oak Ridge, TN, 111 pp

Boyle, J. S., Williamson, D., Cederwall, R., Fiorino, M., Hnilo, J., Olson, J., Phillips, T., Potter, G., and Xie, S. (2005), Diagnosis of Community Atmospheric Model 2 (CAM2) in numerical weather forecast configuration at Atmospheric Radiation Measurement sites, *J. Geophys. Res.*, **110**, D15S15, https://doi.org/10.1029/2004JD005042.

Brient, F., Roehrig, R., and Voldoire, A. (2019) Evaluating Marine Stratocumulus Clouds in the CNRM-CM6-1 Model Using Short-Term Hindcasts, *J. Atmos. Model Dev*., **11**, 127–148, https://doi.org/10.1029/2018MS001461.

Burkhardt, U., Rybka, H., Arka, I., Koehler, M., Seifert, A., Strandgren, J., Sourdeval, O., Reichardt, J.G., Horvath, A., Bugliaro, L. and Quaas, J., (2017). The impact of convection on upper level cloudiness-transpose AMIP simulations using a hierarchy of models. *AGUFM*, pp.A31I-2297.

Clark, A. J., Jirak, I. L., Dembek, S. R., Creager, G. J., Kong, F., Thomas, K. W., Knopfmeier, K. H., Gallo, B. T., Melick, C. J., Xue, M., Brewster, K. A., Jung, Y., Kennedy, A., Dong, X., Markel, J., Gilmore, M., Romine, G. S., Fossell, K. R., Sobash, R. A., Carley, J. R., Ferrier, B. S., Pyle, M., Alexander, C. R., Weiss, S. J., Kain, J. S., Wicker, L. J.,

© Crown copyright 2021, Met Office

Thompson, G., Adams-Selin, R. D., & Imy, D. A. (2018). The Community Leveraged Unified Ensemble (CLUE) in the 2016 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment, *Bull. Amer. Meteor. Soc.*, **99**, 1433-1448. https://doi.org/10.1175/BAMS-D-16-0309.1

Clark A, Jirak, I, Gallo, BT, Dean, A, Knopfmeier, K, Roberts, B, Wicker, L, Krocak M, Skinner, P, Heinselman, P, Wilson, K, Vancil, J, Hoogewind, K, Dahl, N, Creager, G, Jones, T, Gao, J, Wang, Y, Loken, ED, Flora, M, Kerr, C, Yussouf N, Dembek, S, Miller, W, Martin, J, Guerra, J, Matilla, B, Jahn, D, Harrison, D, and Imy, D (2020a) Spring Forecasting Experiment 2020: Program Overview and Operations Plan. *NOAA SPC/NSSL Report,* 36 pp. https://hwt.nssl.noaa.gov/sfe/2020/docs/eval_guide_sfe2020.pdf.

Clark, A, Jirak, I, Gallo, BT, Roberts, B, Dean, A, Knopfmeier K, Wicker, L, Krocak, M, Skinner, P, Heinselman, P, Wilson, K, Vancil, J, Hoogewind, K, Dahl, N, Creager, G, Jones, T, Gao, J, Wang, Y, Loken, ED, Flora, M, Kerr, C, Yussouf, N, Dembek, S, Miller, W, Martin, J, Guerra, J, Matilla, B, Jahn, D, Harrison, D, Imy, D, and Coniglio, M (2020b) Spring Forecasting Experiment 2020: Preliminary Findings and Results. *NOAA SPC/NSSL Report* 77 pp. https://hwt.nssl.noaa.gov/sfe/2020/docs/HWT_SFE_2020_Prelim_Findings_FINAL.pdf.

Clark, P. A., C. E. Halliwell, and D. L. A. Flack, (2021) A Physically-Based Stochastic Boundary-Layer Perturbation Scheme. Part I: Formulation and evaluation in a convection-permitting model. *J. Atmos. Sci.* ***78****,* 727-746. https://doi.org/10.1175/JAS-D-19-0291.1

Flack, D. L. A., S. L. Gray, R. S. Plant, H. W. Lean, and G. C. Craig (2018) Convective-Scale Perturbation Growth across the Spectrum of Convective Regimes. *Mon. Wea. Rev.*, **146**, 387–405, https://doi.org/10.1175/MWR-D-17-0024.1.

Flack, DLA, Gray, SL, Plant, RS. (2019) A simple ensemble approach for more robust process-based sensitivity analysis of case studies in convection-permitting models. *Quart. J. Roy. Meteor. Soc*. 145, 3089– 3101, https://doi.org/10.1002/qj.3606

Flack, DLA, PA Clark, CE Halliwell, NM Roberts, SL Gray, RS Plant, and HW Lean, (2021a*)* A Physically-Based Stochastic Boundary-Layer Perturbation Scheme. Part II: Perturbation

© Crown copyright 2021, Met Office

Growth within a Super Ensemble Framework. *J. Atmos. Sci.**78**,* 747-761. https://doi.org/10.1175/JAS-D-19-0292.1

Flack, D. L. A., Rivière, G., Musat, I., Roehrig, R., Bony, S., Delanoë, J., Cazenave, Q., and Pelon, J. (2021b) Representation by two climate models of the dynamical and diabatic processes involved in the development of an explosively deepening cycle during NAWDEX, *Wea. Clim. Dyn.*, **2**, 233-253. https://doi.org/10.5194/wcd-2-233-2021

Garcia-Moya, J.-A., Callado, A., Escriba, P., Santos, C., Santos-Munoz, D., and Simarro, J. (2011) Predictability of short-range forecasting: a multimodel approach*. Tellus A*, **63**, 550-563. https://doi.org/10.1111/j.1600-0870.2010.00506.x.

Hanley, K.E., Plant, R.S., Stein, T.H.M., Hogan, R.J., Nicol, J.C., Lean, H.W., Halliwell, C. and Clark, P.A. (2015), Mixing-length controls on high-resolution simulations of convective storms*. Quart. J. Roy. Meteor. Soc.*, 141, 272-284. https://doi.org/10.1002/qj.2356.

Karmalkar, A., (2015) Characterization of climate model errors over North America at climate and NWP timescales using the NARCCAP RCM and Transpose-AMIP experiments. *AGUFM*, pp.A21E-0191.

Karmalkar, A. and Bradley, R.S., (2016) Characterization of climate model errors over North America using the NARCCAP RCM, CMIP5-AMIP and Transpose-AMIP experiments. *AGUFM*, pp.A54D-06.

Keat, WJ, Stein, THM, Phaduli, E, Landman, S., Becker, E, Bopape M-J. M., Hanley, KE, Lean, HW, Webster, S. (2019) Convective initiation and storm life cycles in convection-permitting simulations of the Met Office Unified Model over South Africa. *Quart. J. Roy. Meteor. Soc.*, **145**: 1323– 1336. https://doi.org/10.1002/qj.3487.

Keil, C., Heinlein, F. and Craig, G.C. (2014), The convective adjustment time-scale as indicator of predictability of convective precipitation. *Quart.J.Roy. Meteor. Soc.*, **140,** 480-490. https://doi.org/10.1002/qj.2143

Klocke, D. and Rodwell, M. J. (2014) A comparison of two numerical weather prediction methods for diagnosing fast-physics errors in climate models*, Quart. J. Roy. Meteor. Soc.*, **140**, 517–524, https://doi.org/10.1002/qj.2172.

© Crown copyright 2021, Met Office

Kühnlein, C., Keil, C., Craig, G.C. and Gebhardt, C. (2014), The impact of downscaled initial condition perturbations on convective-scale ensemble forecasts of precipitation. *Quart. J. Roy. Meteor. Soc.*, **140**: 1552-1562. https://doi.org/10.1002/qj.2238.

Leoncini, G., Plant, R.S., Gray, S.L. and Clark, P.A. (2010), Perturbation growth at the convective scale for CSIP IOP18. *Quart. J. Roy. Meteor. Soc*., **136**, 653-670. https://doi.org/10.1002/qj.587.

Li, J., Chen, H., Rong, X., Su, J., Xin, Y., Furtado, K., Milton, S., and Li, N. (2018) How Well Can a Climate Model Simulate an Extreme Precipitation Event: A Case Study Using the Transpose-AMIP Experiment, *J. Clim.*, **31**, 6543–6556, https://doi.org/10.1175/JCLI-D17-0801.1.

Ma, H.-Y., Xie, S., Boyle, J. S., Klein, S. A., and Zhang, Y. (2013) Metrics and Diagnostics for Precipitation-Related Processes in Climate Model Short-Range Hindcasts, *J. Climate*, **26**, 1516–1534, https://doi.org/10.1175/JCLI-D-12-00235.1.

Marsigli, C., Montani, A. and Paccagnella, T. (2014), Perturbation of initial and boundary conditions for a limited-area ensemble: multi-model versus single-model approach. *Quart. J. Roy. Meteor. Soc*., **140**: 197-208. https://doi.org/10.1002/qj.2128

Pearson, K. J., Shaffrey, L. C., Methven, J., and Hodges, K. I. (2015) Can a climate model reproduce extreme regional precipitation events over England and Wales?, *Quart. J. Roy. Meteor. Soc.,* **141**, 1466–1472, https://doi.org/10.1002/qj.2428.

Phillips, T. J., Potter, G. L., Williamson, D. L., Cederwall, R. T., Boyle, J. S., Fiorino, M., Hnilo, J. J., Olson, J. G., Xie, S., and Yio, J. J. (2004) Evaluating Parameterizations in General Circulation Models: Climate Simulation Meets Weather Prediction*, Bull. Amer. Meteor. Soc*., **85**, 1903–1916, https://doi.org/10.1175/BAMS-85-12-1903.

Porson, AN, Hagelin, S, Boyd, DFA, Roberts, N.M., North, R., Webster, S., Lo, J. C.-F. (2019) Extreme rainfall sensitivity in convective-scale ensemble modelling over Singapore. *Quart. J. Roy. Meteor Soc*. **145**, 3004– 3022. https://doi.org/10.1002/qj.3601

© Crown copyright 2021, Met Office

Porson, A, Roberts N, Walters, D, Willington, S, Gilbert R, McCabe A, Dow, G, Flack, D, Mittermaier M, Bush, M, Richter, H (2020) HWT 2020, Met Office Internal Report. https://metoffice-my.sharepoint.com/:b:/r/personal/aurore_porson_metoffice_gov_uk/Documents/Documents/PEG_HWTreport_Dec2020.pdf?csf=1&web=1&e=CITdNP

Pscheidt, I, Senf, F, Heinze, R, Deneke, H, Trömel, S, Hohenegger, C. (2019) How organized is deep convection over Germany?. *Quart. J. Roy. Meteor Soc*. **145,** 2366– 2384. https://doi.org/10.1002/qj.3552.

Roff, G. (2015) GA6 Transpose-AMIP: Procedure and preliminary comparison with ACCESS 1.3. *CAWCR Res. Let*. https://accessdev.nci.org.au/trac/raw-attachment/wiki/access/T-AMIP_GA6/2015%20paper%20CAWCR_Res_Let_NUM_Roff_TAMIP_GA6.pdf

Roff, G. and Zhang, H. (2015) Transpose-AMIP experiments for testing the potential impacts of land-surface models on ACCESS NWP. In *Coupled Modelling and Prediction: from weather to climate-abstracts of the ninth CAWCR Workshop 19 October-22 October 2015, Melbourne, Australia* (p. 101).

Selz, T., and Craig, G. C. (2015). Upscale Error Growth in a High-Resolution Simulation of a Summertime Weather Event over Europe, *Mon. Wea. Rev.*, **143**, 813-827. https://doi.org/10.1175/MWR-D-14-00140.1

Sexton, D. M. H., Karmalkar, A. V., Murphy, J. M., Williams, K. D., Boutle, I. A., Morcrette, C. J., Stirling, A. J., and Vosper, S. B. (2019) Finding plausible and diverse variants of a climate model. Part I: establishing the relationship between errors at weather and climate time scales, *Clim. Dyn*., **53**, 989–1022, https://doi.org/10.1007/s00382-019-04625-3.

Stein, T. H. M., R. J. Hogan, P. A. Clark, C. E. Halliwell, K. E. Hanley, H. W. Lean, J. C. Nicol, and R. S. Plant, (2015) The DYMECS Project: A Statistical Approach for the Evaluation of Convective Storms in High-Resolution NWP Models. *Bull. Amer. Meteor. Soc.*, **96**, 939–951, https://doi.org/10.1175/BAMS-D-13-00279.1.

Williams, K. D., Bodas-Salcedo, A., Déqué, M., Fermepin, S., Medeiros, B., Watanabe, M., Jakob, C., Klein, S. A., Senior, C. A., and Williamson, D. L. (2013) The Transpose-AMIP II

© Crown copyright 2021, Met Office

Experiment and Its Application to the Understanding of Southern Ocean Cloud Biases in Climate Models*, J. Clim.*, **26**, 3258–3274, https://doi.org/10.1175/JCLI-D-12-00429.1.

Wilks, D. S. (2011) Statistical methods in the atmospheric sciences. *3rd Edition*. Academic press, 676 pp.

Zhang, F., Snyder, C. and Rotunno, R., (2003). Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60**, 1173-1185. https://doi.org/10.1175/1520-0469(2003)060<1173:EOMCOM>2.0.CO;2.

© Crown copyright 2021, Met Office

© Crown copyright 2021, Met Office

Met Office
FitzRoy Road
Exeter
Devon
EX1 3PB
United Kingdom

© Crown copyright 2021, Met Office