Problems with ISIMIP3bV3 Tmax (tasmaxAdjust May 2021 version) for use with population based research.

Bruno Lemke, Tord Kjellstrom, Chris Freyberg, Matthias Otto.
11 June 2021

Aim
As requested by Stefan Lange: to check the improvements in tasmaxAdjust for Version 3 of ISIMIP3b. This is the version added to the ISIMIP database in May 2021

Executive Summary:
This report finds that there are still significant issues in the May 2021 ISIMIP3b historical tasmaxAdjust grid cell data compared to weather station data and previous baseline datasets. However, these issues do not seem to arise from the process of downscaling and bias correction but seem to arise from the change from using the W5E5 Tmax dataset for bias adjustment instead of the EWEMBI Tmax dataset. Most of the problems arise from coastal cells with less than 50% land and also in a small cluster of cells in Vietnam. This new dataset makes it very difficult to do accurate population based studies at a local level because large populations live in the coastal grid cells. We are wondering whether a land only climate dataset could be produced using EWEMBI as a baseline for bias adjustment of coastal cells rather than W5E5. Other suggestions are made in the report below.

Materials and Methods:

Table 1 Terminology used in this report for ISIMIP3b variables and versions		
Variable	**Our terminology**	**ISIMIP terminology**
Maximum daily temperature (surface, bias corrected)	Tmax	tasmaxAdjust
Average daily temperature (surface, bias corrected)	Tmean	tasAdjust
Minimum daily temperature (surface, bias corrected)	Tmin	tasminAdjust
Average daily relative humidity	RH	hurs
Average daily downwelling shortwave radiation	rsds	rsds
ISIMIP3b models available March 2020	GFDL4V1 etc	GFDL-esm4 etc
ISIMIP3b models available June 2020	GFDL4V2 etc	GFDL-esm4 etc
ISIMIP3b models available May 2021	GFDL4V3 etc	GFDL-esm4 etc

<u>Three methods</u>
Method 1: Ten years of daily ISIMIP3bV3 grid cell data (2001-2010) from the PIK data repository (esg.pik-potsdam.de/search/isimip/) was processed into 10-year monthly data (see table 1) by Chris Freyberg as the daily data was not convenient for our impact studies (heat stress on workers). This was combined with our 0.5x0.5 degree grid based Tmax data of 10 years (2001-2010) that had average Tmax values for weather stations allocated to grid cells, CRU4, CPC (daily gridded data from NOAA) and various versions of ISIMIP generated data: EWEMBI1, W5E5, ISIMIP2b and ISIMIP3bV2. We then compared these various Tmax grid cell datasets. As we had most faith in the weather station data, we looked first at the largest difference between calendar month (10-year average) weather station Tmax data and UKesmV3 (as a representative of ISIMIP3bV3) data in the historic timeframe for grid cells where there were weather stations.

Method 2: Once we had identified problems with the monthly 10-year averaged data by method 1, we used annual monthly CRU data (crudata.uea.ac.uk/cru/data/hrg/) as a grid cell proxy for weather station data and then actual daily HothapsSoft weather station data (available from Climatechip.org) and derived from the NOAA GSOD database (data.noaa.gov/dataset/dataset/global-surface-summary-of-the-day-gsod).

Method 3: Daily studies of Tmax for specific grid cells were carried out using the Panoply Software from NOAA (www.giss.nasa.gov/tools/panoply/download) on the original (unprocessed by us) daily ISIMIP3b data (May 2021 version)

Results:

We discovered two issues: The first issue, "the widespread coastal problem", occurs in many coastal grid cells. This issue affects our human impact calculations badly resulting in a 20% reduced impact in some countries because of large populations in coastal cells. However, if a research group is doing non-land based impact studies, then this "problem" is unlikely to be an issue for them.
The second issue, "the one-off Vietnam problem", showed by far the greatest difference between grid cell and weather station data, but we have only found one major grid cell cluster like this in the global land based cells we investigated.

1. **Widespread coastal problem which is an issue for researchers who focus on population or other land-based impact studies.**

We discovered this problem by looking at cells with one of the greatest difference between the 10-year average of monthly Tmax of weather stations and UKesmV3. For example, the cell that covers Bahrain (26.25, 50.75) is a cell where the 10-year monthly averages of both UKesmV3 and GFDL4V3 have a Tmax in May of 30C while the weather station (GSOD) give a value of Tmax closer to 40C, a difference of 10C. The months of April, June and July also have similar large Tmax differences. What is also interesting about this problem is that EWEMBI, GFDL2b and HadGEM2b (both bias corrected from EWEMBI) seem to be OK with the Bahrain Tmax for May = 41C, but W5E5, GFDL4 and UKesm (both bias corrected from W5E5) have Tmax over 10C lower (at 30C) AND similar to the Tmax of ERA5 (30C). It is our understanding that W5E5 is a combination of EWEMBI over land and W5E5 over sea. Bahrain is in a grid cell that has 20% land and 80% sea (see Figure 1). Of interest is to note if one moves away from land into the Persian Gulf (a few grid cells from Bahrain), the sea temperature is 34C the same as for Bahrain in UKesmV3.

To confirm this 10-year monthly difference we looked at the **daily** difference directly from the downloaded UKesmV3 data (without any processing by us – see Method 3). The left to right transverse across the latitude of Bahrain is shown in Figure 1. Clearly Bahrain has been assigned a sea temperature from ERA5 in data sets bias corrected using W5E5 and neighbouring land cells have been assigned temperatures bias corrected with EWEMBI. Surprisingly, cells to the immediately west and east of the Bahrain cell have approximately the same proportion of land to sea but they have been assigned land temperatures in UKesmV3.

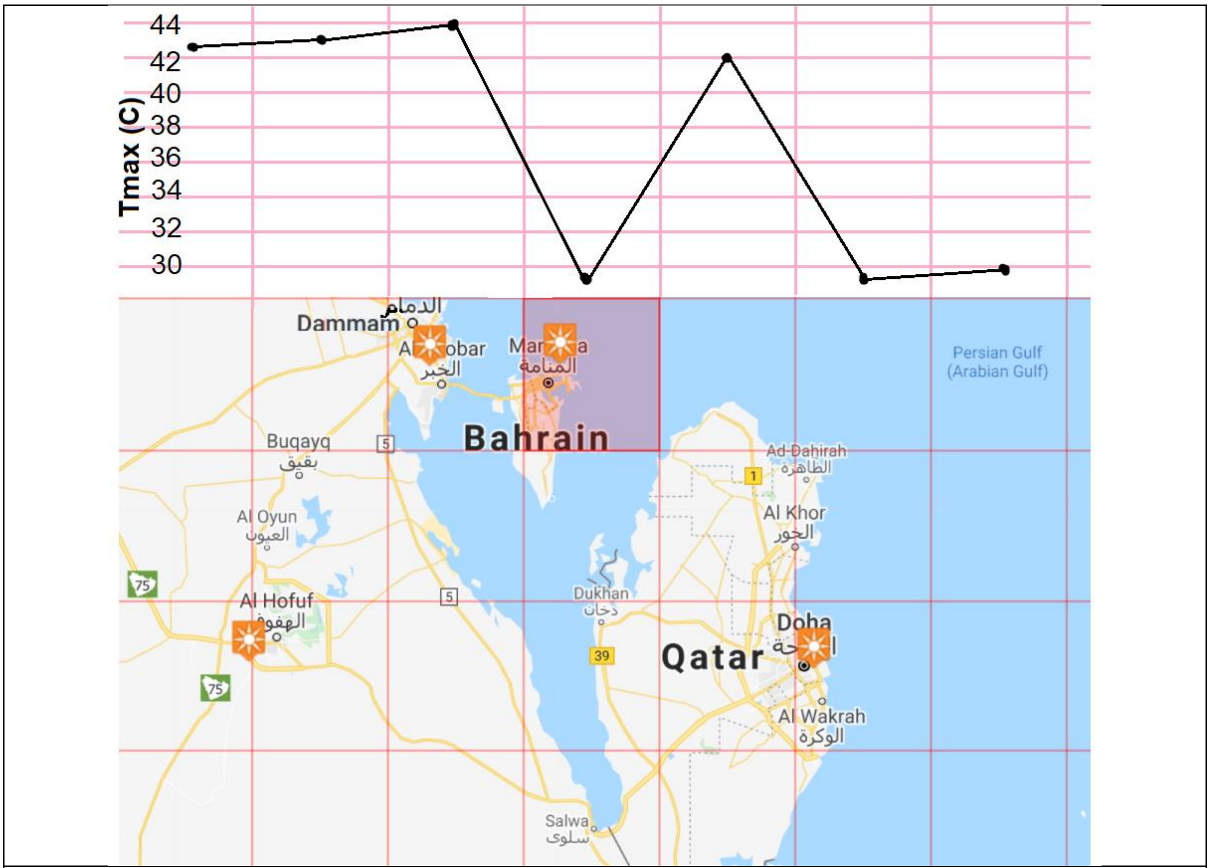


Figure 1. Map from climatechip.org gives the land profile in each 0.5x0.5 degree grid cell around Bahrain (Lat = 26.25, Lon = 50.75 – highlighted). Good quality weather stations are marked by orange sun symbols. The graph immediately above is the bias corrected daily maximum near surface air temperature (Tmax) of Bahrain and neighbouring cells at the same latitude as Bahrain. The date of the daily Tmax data from UKesm4V3 is 3rd May 2003.

This is not an issue limited to a few places.

Without going into the time-consuming studies of daily data, we used our 10-year (2001-2010) Tmax values from our global Tmax data source described in Method 1 to further investigate the geographic spread of this effect. We used the following fingerprint: Grid cells containing a proportion of land and sea and where Tmax from EWEMBI and W5E5 differ. We applied the criterion of a greater than 2C Tmax difference between EWEMBI and W5E5 in order to lift this effect out of the general data scatter (noise) in EWEMBI and W5E5 (see figure 5 later).

EWEMBI does not appear to have this coastal issue until the land proportion in a grid cell is less than 10%. We anticipate that an ocean sharing part of a cell will generally produce a cooling effect in summer but in winter in colder regions, this may result in a warming effect. For some months Tmax from W5E5 was larger than EWEMBI by 2C in some coastal cells. This effect was considerably less than where ocean temperatures were cooler than land temperatures and will be discussed separately in the next section.

Monthly 10-year average Tmax data

We start with cells where the 10-year average of Tmax(W5E5) is lower than Tmax(EWEMBI) by more than 2C for **any month**. We count "cell-months" as the combined number of months for all affected cells. Figure 2 shows that for cells with a land cover of 100% right down to 50%, there are almost no cells where Tmax(W5E5) is lower than Tmax(EWEMBI) by more than 2C. But, as the land percentage falls below 50%, cells numbers where Tmax(W5E5) is lower than Tmax(EWEMBI) by more than 2C rapidly increases to nearly 1400 cell-months for **all months**. Many cells have Tmax(W5E5) lower than Tmax(EWEMBI) for more than one month so Figure 3 shows data for the 909 grid cells where each cell is counted only once no matter how many months their Tmax(W5E5) is lower than Tmax(EWEMBI) by more than 2C. Note that the graphs are NOT accumulated values but the number of cells in every 5% coastal land area bin. The total population in these coastal cells where the percentage of land is less than 50% is 24,769,379 people. Figure 4 shows the distribution of the 454 grid cells where Tmax(W5E5) is 5C lower than Tmax(EWEMBI) and there are 28 cells where Tmax(W5E5) is 10C or more lower than Tmax(EWEMBI). See list of these 28 cells in Appendix 1. The population in the 5C Tmax difference grid cells is 14,310,060 and the population of the 10C Tmax difference grid cells is 1,627,681.

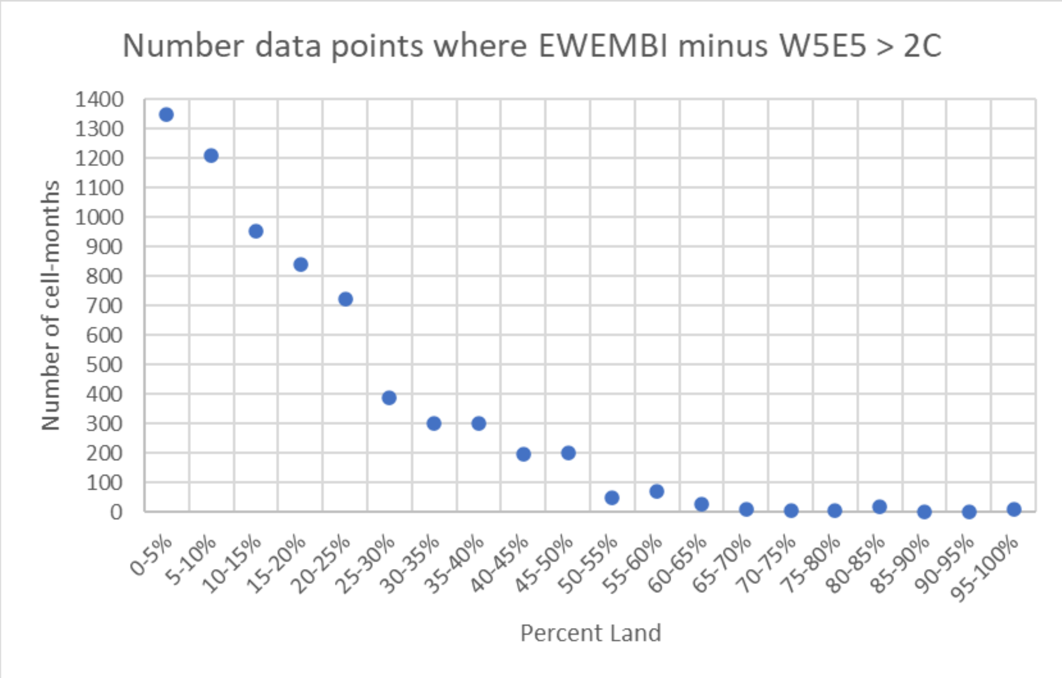


Figure 2. Number of cell-months where Tmax(EWEMBI) minus Tmax(W5E5) for 2001-2010 monthly average for **all months** is greater than 2C. Data NOT accumulative, but actual values in each 5% land area bin. Note that **each month** a grid cell meets this condition has been counted.

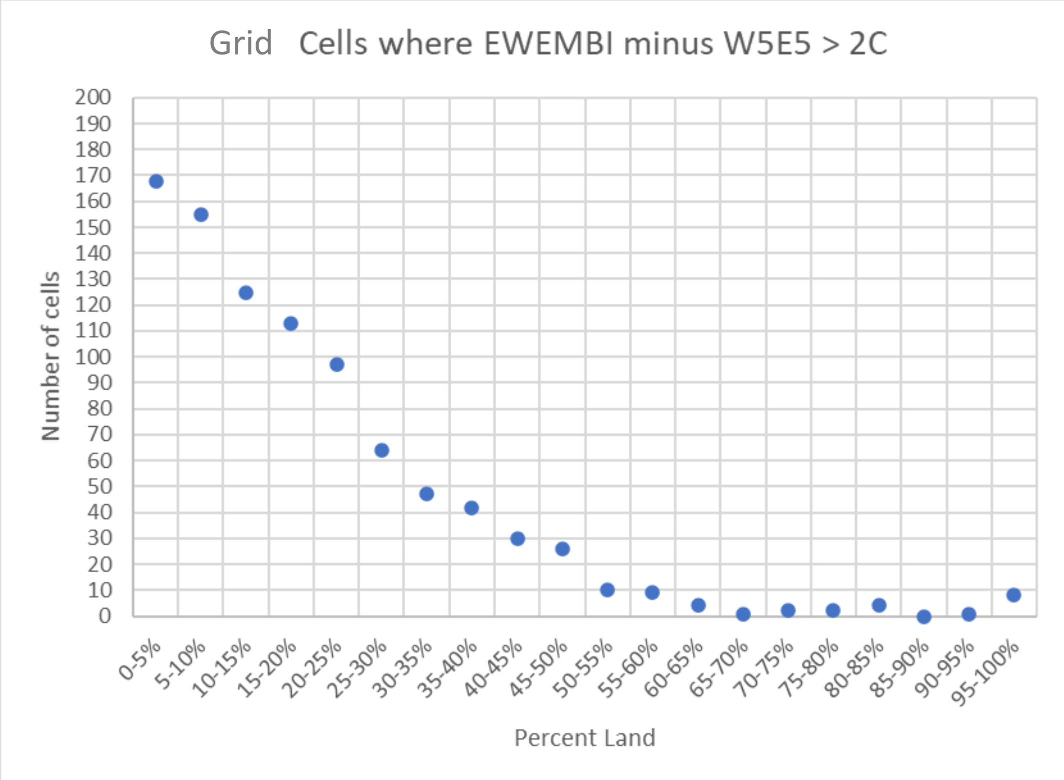


Figure 3. Number of grid cells where Tmax(EWEMBI) minus Tmax(W5E5) for 2001-2010 monthly average is greater than 2C **for at least one month**. Data NOT accumulative, but actual values in each 5% land area bin. Note in this graph each cell is counted only once.

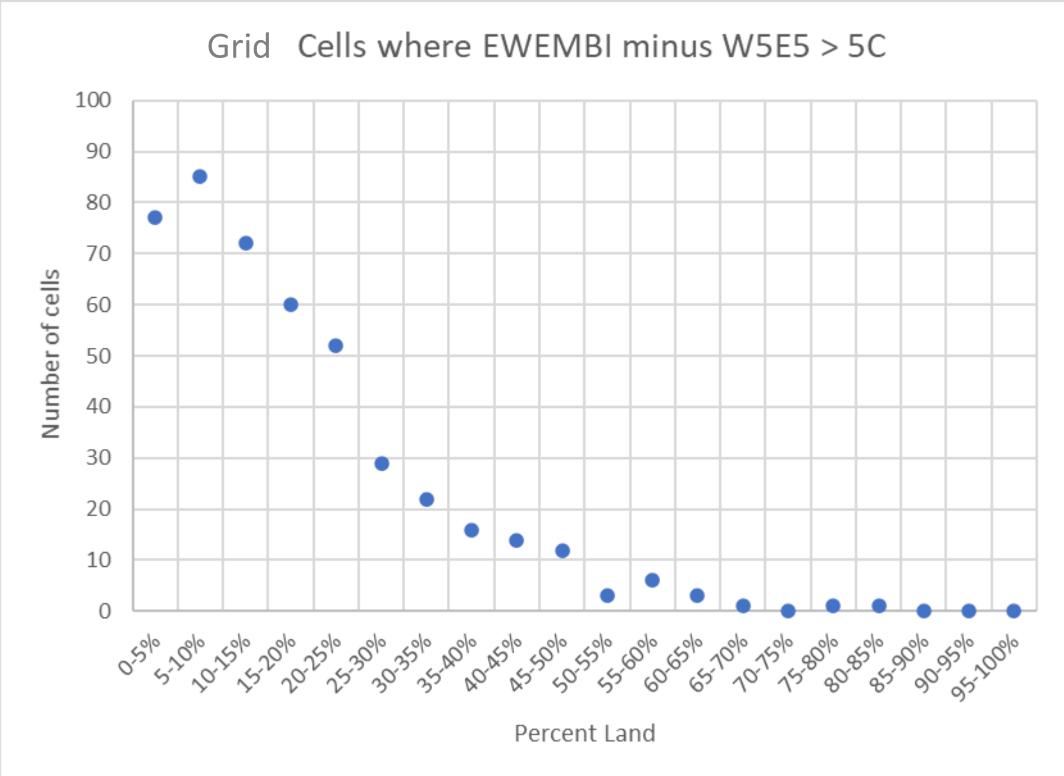


Figure 4. Number of grid cells where Tmax(EWEMBI) minus Tmax(W5E5) for 2001-2010 monthly average is greater than 5C for at **least one month**. Data NOT accumulative, but actual values in each 5% land area bin. Each cell only counted once even if it exceeds the criteria in many months

Comparison with CRU4

As the number of weather stations in coastal grid cells with less than 50% land is low (556 out of 60,591 cells over land) and sometimes weather stations have incomplete data or have errors (see discussion), we used CRU4 data as a proxy for weather stations to highlight the coastal effect of W5E5 (see method 2). CRU gridded data is based solely on weather stations and includes quality control checks (Ian Harris, Timothy J. Osborn, Phil Jones & David Lister www.nature.com/articles/s41597-020-0453-3). Figure 5 compares CRU4 with W5E5. The +/- 2C scatter is apparent for all grid cells but this scatter increases greatly and with a negative bias when the land occupies less than 50% of a grid cell.

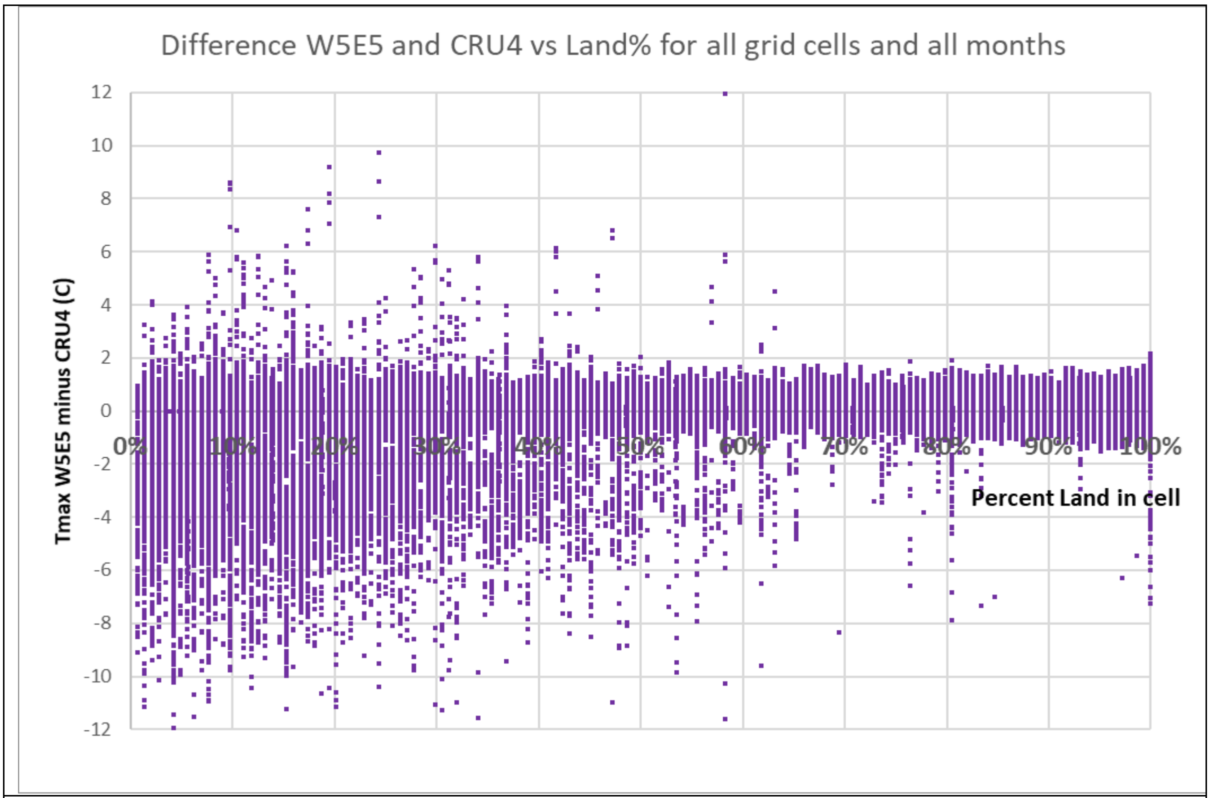


Figure 5. Scatter graph for all land grid cells and all months 2001-2010 average for the Tmax difference between W5E5 and CRU4

When Tmax(W5E5) is greater than Tmax(CRU) for coastal cells

Figure 5 confirms the results from the earlier data that W5E5 has coastal cells with significantly cooler Tmax than CRU (as a proxy for weather stations). However it can be seen that some W5E5 grid cells also have coastal cells with a much higher Tmax than CRU (or EWEMBI). Oceans are generally cooler than the maximum land temperature, but in colder climates oceans temperature can be higher than **maximum** land temperatures in some months.

It is clear from Figure 6 that when the land grid cell temperatures are high, the effect of the more than 50% ocean in that cell is cooling while if the grid cell is cold (winter months) there can also be significant warming by the ocean for those grid cells. While a number of these cooler coastal cells were in polar regions there were 81 unique coastal cells outside the poles that had Tmax(W5E5) greater than Tmax(EWEMBI) by more than 2C. Only 30 of these cells (out of 6693 cells) had a difference that was more than 3C. The total population of these cells was only 175,541. See list of these cells in appendix 2.

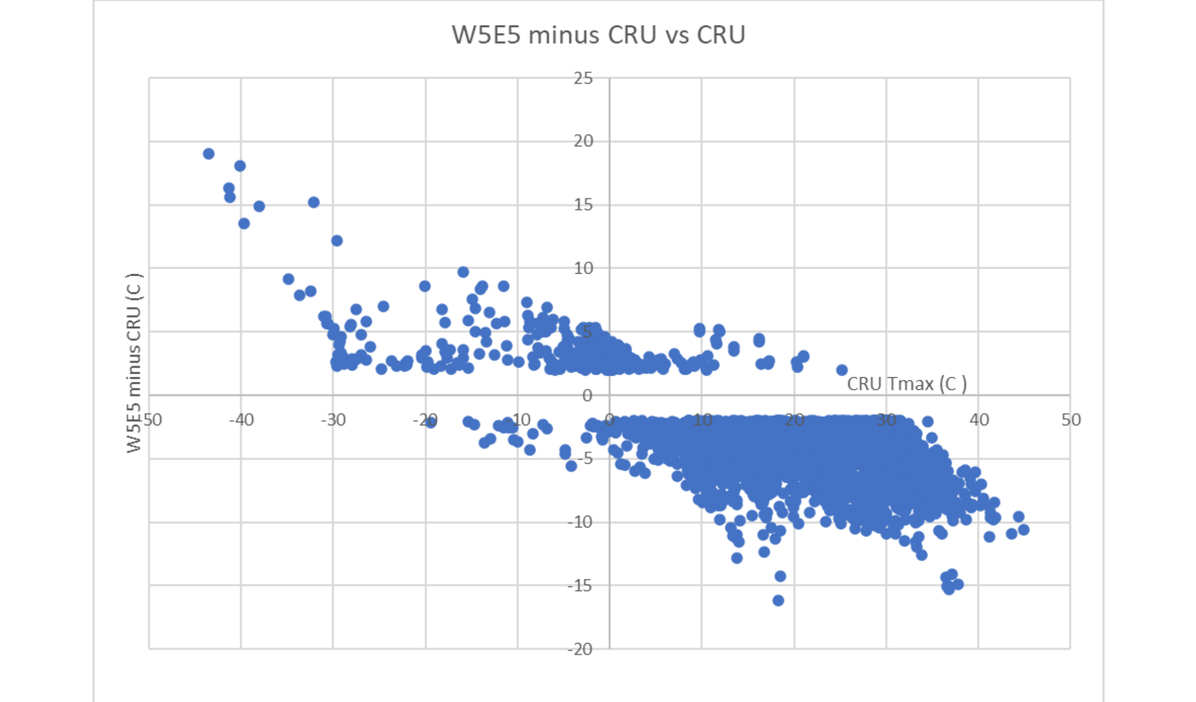


Figure 6. Scatter graph for all land grid cells and all months 2001-2010 average for the Tmax difference between Tmax(W5E5) minus Tmax(CRU4) vs Tmax(CRU4) (as a grid cell proxy for weather station data). Only differences of more than 2C between W5E5 and CRU4 are plotted to remove the general scatter in Tmax difference in all cells between these 2 models (Figure 5)

2. **The "one off Vietnam" problem:**

The weather station in the city of Quy Nhon in Vietnam (grid cell coordinates 13.73, 109.25) has its maximum daily temperature, as recorded in GSOD, in March 2001-2010 ranging from 28.5C to 31C. For that grid cell, the CRU4 March average from 2001 to 2010 was 28.6C, the CPC (NOAA grid cell) was 29.8C, GWSP (from ISIMIP2a) was 31.2C and ERA5 was 26.6C. However, the March 2001-2010 average of Tmax for UKesm (May 2021 version) was 19.8C, some 10C lower than actual. There were similar low Tmax values for UKesmV2 (19.0C), HadGEM2b (22.3C), GFDL4V3 (19.3C) and GFDL2b (22.1C). The EWEMBI data for that grid cell for the March (2001-2010) was 19.1C, W5E5 was 19.0C. Inland grid cells 0.5 degrees to the north and up to 1 degree to the south from (13.75, 109.25) also have the same problem but the difference is nearer 5C rather than 10C. Surprisingly, while the CRU Tmax closely matches the weather station temperature, the CRU Tmean in these cells is considerably less in March than for the same cells in February and April. We checked the Monsoon season for Quy Nhon and March is one of the driest month with the wet monsoons occurring between September and December.

Annual and Daily Data

CRU annual data (rather than 10-year averages) for March (from Climatechip.org) gives the March average Tmax over each of the 10 years ranging from 28-29C for the Quy Nhon grid cell. To study daily Tmax we used Panoply (see Method 3) on the May 2021 UKesmV3 data. The Quy Nhon grid cell is 17C lower than nearby inland grid cells and 9C lower than the grid cells over the ocean (Figure 7). As we scrolled through the days from the beginning of 2004, when we hit the start of March there is a jump where the Tmax goes down by about 10C and then quickly goes back to what is expected at the end of March. This strange dip is in March for every year from 2001-2010 with 2004 dropping to 13.4C but 2001, 2002, 2008, 2009 dropping to 15C (Figure 8). Note that only the years 2001-2010 were studied.

All cells where the land area is greater than 50% (see later) and where the difference between W5E5 and EWEMBI was less than 1C and the difference between W5E5/EWEMBI and CRU is more than 3C (similar to the "one-off Vietnam" effect) were all located around the Quy Nhoy in Vietnam plus a small cluster the North West tip of Iceland. The 24 Vietnamese cells falling into this category had a total population of 9,345,357 while the 5 Icelandic cells had a population of 2998. While all months were tested, this effect in Vietnam and Iceland occurred only in March. This gives a clear fingerprint to this particular effect that we found nowhere else.

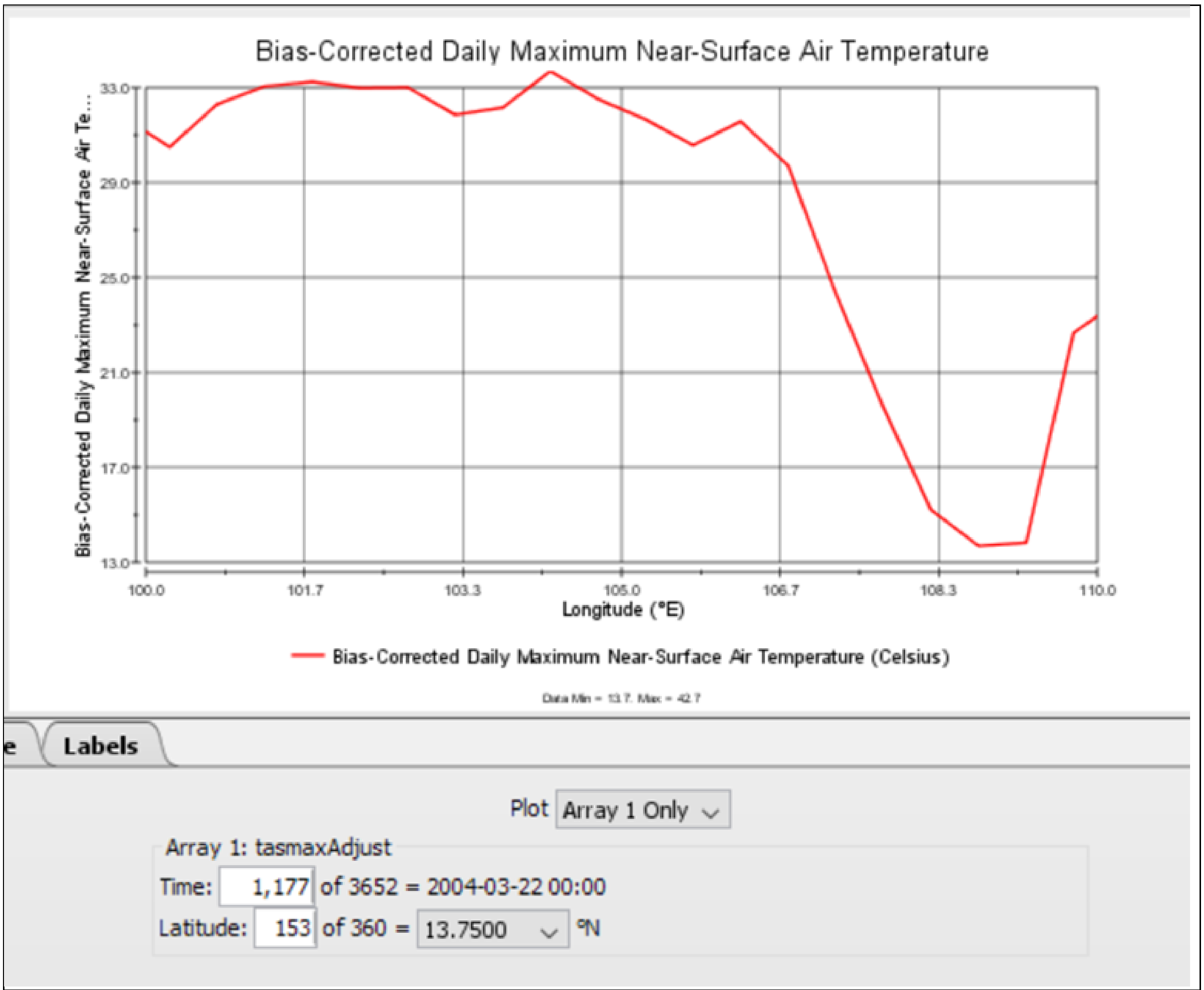


Figure 7. UKesmV3 Tmax at latitude of the Quy Nhon grid cell along with neighbouring longitudinal grid cells, for 22 March 2004. The Tmax of the Pacific Ocean is shown in the last two grid cells on the right, the Quy Nhon grid cell is where the Tmax drops to 13.7C and the cells to the left of that go inland from Quy Nhon. Each graph kink is the centre of an 0.5 grid cell.

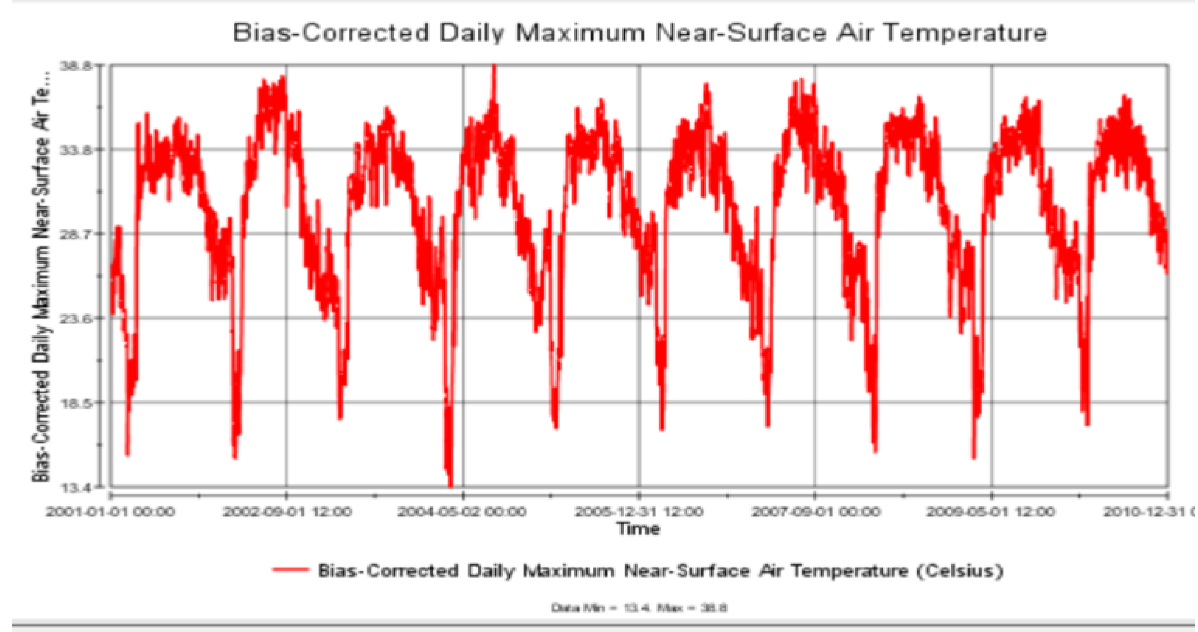


Figure 8. The Quy Nhon UKesmV3 Grid cell with Tmax plotted for every day 2001-2010. Showing the sudden drop in March each year.

Discussion

1. The Widespread Coastal Problem

EWEMBI is primarily based on WATCH data derived from weather stations while W5E5 is a combination of land-based WATCH data and ocean based ERA5 data (dataservices.gfz-potsdam.de/pik/showshort.php?id=escidoc:4855898). This can cause significant problems in coastal cells where many people live because, once a cell is over 50% water it seems that in W5E5 that cell temperatures are biased toward water temperature while in EWEMBI it is biased toward the land temperature. EWEMBI seems to retain its land bias right down to 10% land area in that cell.

This issue with the coastal cells could be overcome by sticking with ISIMIP2b rather than progressing to ISIMIP3b. It would be great if PIK produced a land only dataset that completely ignored the ocean temperature – ie employing WATCH data as a bias-correcting baseline for coastal cells or at least cells that has more than 10% land. This coastal problem is also an issue in ERA5 data and one of our team (Chris Freyberg – see appendix 4) has produced an algorithm that corrects this problem in ERA5. ERA5 only extends to 2020 and so is not useful for impact calculations to the end of the century. It might be possible to do a Field Change adjustment to ISIMIP3b coastal cells using EWEMBI as a baseline only for coastal cells.
If we were to use ISIMIP3bV3, we will need to be explore some method of correction because without that, for countries with large coastlines such as Philippines and Indonesia, the effect on our measured health impact (work hours lost due to heat) has been calculated as being up to a 20% less for those countries!

2. The One-off Vietnam Problem

As these strange results seems also to be in the CRU4 database (with Tmean rather than Tmax) we believe that its foundation is in CRU4 and CRU3 Tmax data and there is no concern about the actual processing done by PIK. Other than manually correcting this problem for two dozen cells (see appendix 3), there seems no easy fix. This area is of interest to us in our impact studies in South East Asia so if this cannot be corrected by PIK we will need to do further research on how we can compensate for this.

Conclusion.
There are some issues relating to maximum temperature with the new version of ISIMIP3b that do not come from the downscaling or bias correction but appear to come from baseline data used in the bias correction. These issues are significant for land-based impact studies (eg population-dependent studies). It would be great if ISIMIP could produce both a land-based data set (with appropriate bias-correction) as well as the current global data set (with its equal land-ocean grid cell bias correction).

Appendix 1:
Grid cells where Tmax(W5E5) is more than 10C lower than Tmax(EWEMBI) – the most extreme cells **for the widespread coastal problem results.**

latitude	longitude	Country ISO3	W5E5 alt (m)	Land Percent	Total Population	Tmax(EWEMBI) 2001-10	Tmax(W5E5) 2001-10
-43.25	-64.25	ARG	5.6	5%	2104	29.63	19.04
-32.75	133.75	AUS	1.7	4%	17	31.48	21.14
-23.25	14.25	NAM	21.8	15%	32	28.08	17.95
14.75	-17.75	SEN	0.3	1%	117367	32.99	22.26
16.75	42.25	SAU	2.1	28%	26708	42.82	32.21
22.75	-110.25	MEX	12.0	4%	1473	35.32	25.31
26.25	50.75	BHR	0.9	20%	1135086	45.16	34.29
27.75	-115.25	MEX	4.0	8%	91	30.92	19.04
30.25	-116.25	MEX	7.2	4%	364	29.41	18.61
33.25	-118.25	USA	11.5	12%	4144	29.56	18.78
43.75	-68.75	USA	0.9	5%	98	26.88	16.79
44.75	-66.25	CAN	6.7	3%	104	25.73	14.89
56.75	-79.25	CAN	3.5	16%	406	17.93	7.31
69.75	170.25	RUS	25.7	32%	37	16.53	5.70
73.25	129.25	RUS	0.2	19%	5	16.34	4.43
73.75	136.25	RUS	1.9	15%	7	12.97	2.48
74.25	135.75	RUS	3.3	30%	17	12.65	2.29
74.25	140.25	RUS	2.1	34%	23	18.20	4.30
75.75	135.75	RUS	5.7	47%	29	17.87	2.07
16.25	-16.75	SEN	0.3	6%	2407	37.59	21.62
23.25	68.25	IND	2.1	7%	4213	37.87	27.75
25.25	-109.25	MEX	1.6	1%	1234	34.59	22.40
12.25	52.25	YEM	3.1	13%	2472	36.10	26.06
15.75	42.25	YEM	0.4	10%	4553	42.18	31.68
22.75	-109.75	MEX	42.4	15%	7543	34.21	23.72
41.25	-69.75	USA	0.6	2%	594	25.63	15.42
54.25	137.75	RUS	26.9	31%	6	17.94	6.74
54.75	137.25	RUS	39.0	24%	6	17.30	7.06

Appendix 2

Most extreme grid cells (excluding polar regions) where Tmax(W5E5) is more than 3C **higher** than Tmax(CRU) in widespread coastal problem. Note that there are NO cells with Tmax(W5E5) more than 10C **higher** than Tmax(CRU) so this is much less of a problem than that shown in Appendix 1 for when Tmax(W5E5) is more than 10C **lower** than Tmax(CRU).

latitude	longitude	Country ISO3	W5E5 alt(m)	Percent Land	Total Population	Tmax(W5E5) 2001-10	Tmax(CRU4) 2001-10
30.25	130.25	JPN	26.85	16%	6134	15.01	9.77
30.25	130.75	JPN	48.11	28%	16876	14.84	9.79
48.25	153.25	RUS	4.47	10%	39	-1.01	-4.81
49.75	154.75	RUS	9.20	11%	84	-1.18	-6.77
50.25	156.25	RUS	22.37	4%	24	-1.54	-4.71
51.75	179.75	USA	12.26	15%	26	2.30	-1.41
57.75	10.75	DNK	2.12	6%	2567	3.83	0.27
58.25	-151.75	USA	12.48	9%	13	2.81	-0.90
58.75	10.75	SWE	2.43	2%	60	3.13	-0.89
58.75	152.25	RUS	56.77	10%	13	-12.11	-17.84
58.75	152.75	RUS	50.16	17%	22	-11.45	-18.24
59.25	-150.25	USA	59.26	11%	22	2.58	-2.56
59.25	-139.75	USA	-1.54	8%	10	3.20	-0.57
59.25	155.25	RUS	42.90	24%	31	-11.4	-20.04
59.75	-147.75	USA	32.49	34%	262	3.33	-0.74
59.75	165.25	RUS	5.80	13%	11	-9.52	-14.55
59.75	170.25	RUS	18.47	10%	7	-5.21	-13.81
51.75	178.75	USA	7.06	23%	27	2.47	-0.81
52.25	177.75	USA	6.95	11%	16	2.19	-1.94
38.25	24.75	GRC	35.31	8%	2575	13.78	10.65
42.25	33.75	TUR	67.07	1%	97	10.22	6.98
45.75	142.25	RUS	10.41	6%	301	-0.80	-3.88
52.75	173.25	USA	30.28	33%	48	2.09	-1.17
54.25	142.25	RUS	14.14	14%	83	-8.55	-13.47
54.75	137.25	RUS	38.97	24%	6	-13.75	-17.34
57.25	10.75	DNK	4.74	17%	7413	4.66	1.46
57.25	11.25	DNK	-0.19	8%	1285	4.62	1.36
59.25	149.25	RUS	13.42	12%	7	-9.46	-15.32
59.25	150.75	RUS	26.11	25%	10944	-9.08	-13.33
52.75	172.75	USA	49.37	43%	79	0.46	-3.19

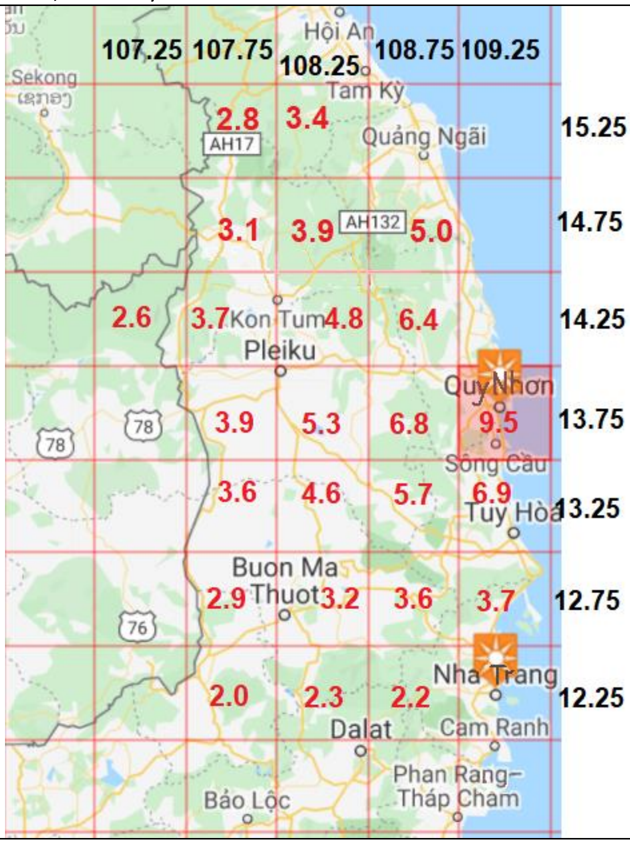
Appendix 3a:

The Tmax issue in Vietnam in March for W5E5 was not apparent in the CRU data, but the same grid cells had a Tmean issue in March for CRU data. These grid cells are identified below where the average Tmean in March was more than 2C lower than the average of February and April Tmean values. All cells are in Vietnam (see appendix 3b).

Lati-tude	Longi-tude	Populat-ion	Alti-tude	CRU 3.24	W5E5	CRU (Feb-Mar)	W5E5 (Feb-Mar)	CRU (Apr-Mar)	W5E5 (Apr-Mar)
11.75	107.75	336998	681	23.22	23.18	0.08	-0.15	2.22	2.10
11.75	108.25	604773	1018	21.42	21.39	0.39	0.14	2.42	2.27
11.75	108.75	459951	590	22.93	22.85	0.10	-0.19	2.28	2.16
12.25	107.25	44054	616	24.13	24.12	0.20	-0.02	2.55	2.38
12.25	107.75	167016	759	22.75	22.75	0.59	0.39	2.81	2.66
12.25	108.25	163942	915	20.97	20.93	0.99	0.74	3.16	3.05
12.25	108.75	82808	889	20.19	20.08	1.19	0.93	3.21	3.17
12.75	107.75	272741	321	23.38	23.35	1.30	1.11	3.76	3.69
12.75	108.25	1123810	527	21.60	21.52	1.89	1.69	4.34	4.28
12.75	108.75	186497	483	20.72	20.58	2.49	2.28	4.91	4.82
12.75	109.25	455719	230	21.72	21.53	2.71	2.45	5.28	5.21
13.25	107.75	53694	198	23.53	23.47	1.91	1.71	4.64	4.53
13.25	108.25	280844	494	20.14	19.98	3.19	3.02	5.71	5.65
13.25	108.75	125788	334	19.13	18.96	4.52	4.29	7.22	7.13
13.25	109.25	606946	109	19.16	18.93	5.61	5.32	8.35	8.32
13.75	107.75	230073	371	21.98	21.80	2.29	2.18	5.00	5.00
13.75	108.25	355023	493	19.37	19.19	3.91	3.72	6.59	6.55
13.75	108.75	251730	444	17.28	17.06	5.62	5.38	8.28	8.21
13.75	109.25	926576	105	16.11	15.83	8.29	8.10	11.32	11.34
14.25	107.25	9267	457	23.67	23.49	0.93	0.74	3.73	3.70
14.25	107.75	217289	601	20.20	20.02	2.12	1.97	4.81	4.73
14.25	108.25	200503	862	17.91	17.67	3.32	3.16	5.88	5.83
14.25	108.75	267750	463	17.63	17.39	5.20	5.03	8.05	8.00
14.75	107.75	124403	871	19.53	19.31	1.52	1.37	4.28	4.23
14.75	108.25	64580	1050	17.95	17.68	2.52	2.35	5.27	5.22
14.75	108.75	417386	368	18.87	18.61	3.81	3.70	6.71	6.68
15.25	108.25	189730	485	20.39	20.14	1.42	1.32	4.44	4.36
15.25	108.75	1062845	86	21.91	21.66	2.21	2.12	5.50	5.43

Appendix 3b

0.5x0.5 grid cells in Vietnam where the March average of Tmean differed by more than 2C from the mean of the February and April monthly average. The difference in ^{o}C is shown in red in those grid cells where the March value was more than 2C lower than the average of February and April. Active weather stations are marked in orange. Longitude and latitude numbers for each grid cell are shown in black along the top and down the right of the figure. The weather station (at Quy Nhon) used to confirm the issue of very low Tmax values for these cells in W5E5 is shown in the pink highlighted cell (13.75, 109.25).



Appendix 4

Brief Summary of the process to correct coastal cell temperatures so that grid cells with less than 50% land retained close to land temperatures rather than an average of sea and land temperatures.

Correction of the daily maximum temperatures of coastal cell in ERA5 "reanalysis-single-levels" dataset at 0.5x0.5 degree resolution. **By Chris Freyberg June 10 2021**

The correction algorithm was developed on the assumption that the underlying model of 2m air temperature is a more-or-less spherical curve whose minimum wavelength is 2 degrees or more and that extracted datasets consist of evenly spaced points on that curve. Such curves will tend to smooth over "discontinuities" that exist at a scale of a few kilometres in climate variables, for example in maximum or minimum 2m air temperatures as one transits from 5km inland of a coast to 5km out to sea. Because ERA5 datasets are geographic point values, it is easiest to explain the algorithm in those terms. The algorithm proceeds as follows:

Locate a set of neighbourhood point values that surround the point of interest (POI) and are more or less equidistant from it and from each other, say the vertices of a (spherical) square. Rank the landfractions of those neighbourhood points. Using some arbitrary but generally applicable cut-off value of land-fraction, say 0.4, choose the subset of points (including the POI) whose land-fraction is greater than that cut-off and calculate the weighted (by land fraction) mean of the climate values at those points to be the corrected climate value at the POI. If none of the set of neighbourhood points nor the POI are associated with a land fraction greater than the cut-off value, use the climate value of the point with the highest land fraction.

For ERA5 and the grid we use, radiating out from the 0.25N,0.25E centre at 0.5 degree spacing, I have used the vertices of each cell as the neighbourhood set of points – that is, for the cell centred at (0.25N, 0.25E) I use (0N,0E), (0N,0.5E), (0.5N,0.5E), (0.5N,0E). I am aided in this by the pretty flexible dataset extraction process made available for ERA5.

Were I to apply this to an ISIMIP dataset, I would use as the neighbourhood set the eight cells adjacent to the cell of interest.