Exploring objective climate classification for the Himalayan arc and adjacent regions using gridded data sources

N. Forsythe[1], S. Blenkinsop[1] and H.J. Fowler[1]

Affiliation:

[1] Centre for Earth Systems Engineering Research (CESER), School of Civil Engineering and Geosciences, Newcastle University, Newcastle, United Kingdom

Corresponding Author contact details:

email: nathan.forsythe@ncl.ac.uk

telephone (office): +44 191 208 6875

telephone (mobile): +44 753 189 3995

Abstract:

A three-step climate classification was applied to a spatial domain covering the Himalayan arc and adjacent plains regions using input data from four global meteorological reanalyses. Input variables were selected based on an understanding of the climatic drivers of regional water resource variability and crop yields. Principal components analysis (PCA) of those variables and k-means clustering on the PCA outputs revealed a reanalysis ensemble consensus for eight macro-climate zones. Spatial statistics of input variables for each zone revealed consistent, distinct climatologies. This climate classification approach has potential both for enhancing assessment of climatic influences on water resources and food security as well as for characterising the skill and bias of gridded datasets, both meteorological reanalyses and climate models, for reproducing sub-regional climatologies. Through their spatial descriptors (area, geographic centroid, elevation mean range), climate classifications also provide metrics, beyond simple changes in individual variables, with which to assess the magnitude of projected climate change. Such sophisticated metrics are of particular interest for regions, including mountainous areas, where natural and anthropogenic systems are expected to be sensitive to incremental climate shifts.

[1] Introduction

The first objective, quantitative systems for global climate classification were developed in the early 20th century by integrating climate data to delineate zones of coherent vegetation type or eco-region (Belda et al., 2014). By distilling information from multiple climate variables which affect vegetation typology, climatic classifications can provide a framework for understanding natural resource systems (Elguindi et al., 2013). By focusing specifically on climate variables which govern river flows and crop growth, derived climate classifications can also yield insight into the dependency of agricultural production on water resources. However, the bulk of recent literature (e.g. Chen and Chen, 2013; Mahlstein et al., 2013; Zhang and Yan, 2014) is global in scope. In this study we focus for the first time on a specific classification for the Himalayan arc and adjacent regions, concentrating on climate types relevant to the spatial domain and time period of interest.

The Himalayan arc and Tibetan Plateau give rise to river systems which sustain populations numbering in the hundreds of millions (Immerzeel et al., 2010). To derive climate classifications for this region we focus on climate variables which control the hydrological regimes of catchments with mountainous headwaters, and hence with substantial runoff contributions from snow and glacial melt, as well crop yields. Our precise study area encompasses the Indus, Ganges and Brahmaputra basins and is shown in Figure 1. The topographic contrast is stark between the high elevation areas of the Himalayan arc and Tibetan plateau, and adjacent lowlands of the Indo-Gangetic plains and deserts of Central Asia. Another striking feature of Figure 1 is the extent of area under irrigation in South Asia. The crops produced by these irrigated surfaces are crucial to the food security of Pakistan, India, Bangladesh and beyond (de Fraiture and Wichelns, 2010). Archer et al. (2010) point out that the semi-arid plains of the Lower Indus had only marginal (rainfed) agricultural viability until the development of irrigation infrastructure. Irrigation demand in the Lower Indus is supplied by run-off from the Hindu Kush, Karakoram and Western Himalaya. Thus holistic understanding of regional food security depends upon characterisation of the spatial as well as climatological differences of these hydrologically-connected sub-regions. Furthermore, it is possible that these sub-regions will experience distinct trajectories of change in the coming decades. Differential rates, or even signs, of change could substantially alter the regional balance of irrigation water supply and demand. The climate classification approach offers a framework within which to evaluate such water balance scenarios.

Global meteorological reanalyses provide coherent syntheses of atmospheric states including radiative and mass flux exchanges with the sea or land surface. In this paper we compare the climatologies described for the study area from four reanalyses – JRA-55 (Ebita et al., 2011), ERA-Interim (Dee et al., 2011), NASA MERRA (Rienecker et al., 2011) and NCEP CFSR (Saha et al.,

2011) – which encompass the recent decades rich in data from both ground-based and satellite-borne instruments. In assessing climate classifications derived from each reanalysis we are not only interested in how the climatically-defined zones relate to water resource supply – mountainous headwaters – and demand – irrigated plains – areas, but also in how the classifications derived from individual reanalyses relate to each other. These inter-comparisons establish a methodology for evaluating gridded datasets, including global and regional climate simulations (Elguindi et al., 2014) as well as reanalyses. Comparisons can be made not only between different models but also between different time periods ("time-slices"), for either historical datasets (Belda et al., 2014; Chen and Chen, 2013) or simulations by climate models (Mahlstein et al., 2013). Temporal changes in derived climate zones can be assessed in terms of both projected spatial changes (areal extent, elevation range, etc.) and of projected climatic changes (mean, annual range, etc.) in the individual climate variables used to create the classification.

[2] Data and Methods

[2.1] Reanalysis datasets

Reanalyses are generally conducted by institutions responsible for meteorological forecasting and are undertaken in part to assess the performance forecasting models and the data assimilation systems which support them (Uppala et al., 2005). The resulting coherent multi-decadal syntheses of climate conditions, however, are of substantial utility to a much broader spectrum of natural scientists. In this study we draw upon data from four reanalyses produced by agencies from diverse geographic regions. Characteristics of the reanalyses used in this study are provided in Table 1 and differ in both spatial and temporal resolutions. Given the forecast-driven nature of reanalyses it is common for time-steps to be organised in 6-hour synoptic forecasting time windows. The NASA MERRA dataset is distinct in that the default time-step is hourly. In all cases daily means were calculated as the mean of the available sub-daily time-steps. Daily maximum and minimum were taken as the highest and lowest values respectively amongst the sub-daily time-steps unless reported specifically as was the case for NCEP CFSR. Diurnal range was calculated as maximum minus minimum. In order to make extracted climatic values as comparable as possible, a common reference period, 1980 to 2009, available from each of the reanalyses, was selected for this study. However, comparability of the results was still limited by differing spatial resolutions of the reanalyses as both temperature and precipitation are greatly influenced by topography in mountainous regions (Immerzeel et al., 2012). The fidelity with which each reanalysis reproduces the topography of the study area is limited by its spatial resolution. For this reason, the JRA-55 – 1.25 x 1.25 decimal degree resolution – dataset is expected to be handicapped compared to NCEP

CFSR – 0.50 x 0.50 decimal degree resolution –dataset. Nevertheless, other elements, including efficacy of data assimilation and realism of land-surface process algorithms, are also expected to play substantial roles in determining reanalysis skill.

[2.2] Selection of climate variables governing water resources and food security

The utility of a climate classification depends on the extent to which it reflects the climatic constraints which govern physical processes of interest. If, for example, geochemical processes such as pollutant mobilisation are an overwhelming concern, sensitivity studies can be conducted to identify the key climatic factors involved (e.g. Nolan et al., 2008). In this paper the processes of interest are river flows from mountainous headwaters and agricultural production, both of which depend upon inputs of mass (precipitation) and energy (ambient temperature and incoming radiation). From a simulation standpoint, common approaches for modelling both meltwater generation from seasonal snowpack and glaciers (Ragettli et al., 2013) as well as crop yields (Baigorria et al., 2007; Kar et al., 2014) require both air temperature and incoming radiation in addition to precipitation as input data. Furthermore, moisture exchanges from the land surface and atmosphere depend upon the latter's vapour pressure deficit which is commonly expressed as relative humidity. Whilst these parameters can be observed directly, the diurnal temperature range (DTR) also acts as an effective proxy for ambient moisture conditions (Easterling et al., 1997).

In establishing the methodology used here, we favoured reanalysis variables with the simplest relationship to commonly observed parameters at ground-based stations. Hence, T_{avg} (mean temperature) and DTR – which together describe the diurnal temperature cycle and can be calculated at stations recording solely~~both calculated from~~ T_{max} (maximum temperature) and T_{min} (minimum temperature) – along with precipitation were selected as governing variables. An exception to this principle was made in selecting net incoming shortwave radiation (SW_{net}) at the ground surface as a governing variable due to the importance of seasonal snow-cover in the hydrological regimes of major Himalayan and Tibetan river systems. SW_{net} can be observed at standard manned meteorological stations and automatic weather station (AWS) units if they are equipped with radiometers, but is also indirectly available from remote sensing via albedo and cloud climatology. It was largely for the linkage between SW_{net} and snow cover via albedo that the former was selected as key variable. Specifically, land surfaces with full snow cover have a much higher albedo than "bare ground" and albedo evolves during snowpack accumulation and ablation when snow cover is partial. Albedo in turn modulates net shortwave absorption from incoming solar radiation at the surface. Thus net shortwave radiation can serve as a proxy for snow cover. The linkage between SW_{net} and cloud cover is also useful as the latter is an indicator of large-scale weather system – mid-latitude westerly or tropical monsoon – influence. Cloud cover influences SW_{net} by modulating the amount of incoming shortwave radiation reaching the surface. In the absence of snow cover, suppression of SW_{net} in summer months over South Asia is likely due to monsoonal activity while suppression in other months suggests mid-latitude westerly disturbances. Table 2 lists the governing variables selected for this study, including the seasonal aggregates of interest, and summarises their physical significance.

Prior to derivation of climate classifications, a comparison of the climatologies from the individual reanalyses provide a context within which differences can be interpreted. To establish a common framework, the "native" resolution data from each reanalysis was regridded (sub-divided) to a common 0.25 * 0.25 decimal degree spatial resolution. Ensemble means were calculated, by grid cell, from the simple averages of the four reanalyses. There was no weighting applied from any metric of skill or confidence, nor were any corrections made to account for differences between "native" orography and estimated surface elevation of the target common grid cell. This approach was taken in the absence of detailed information on likely biases by the reanalyses in the variables of interest. Once the ensemble mean had been calculated, normalised differences, i.e. individual reanalysis value minus ensemble mean, were calculated to facilitate comparisons of individual climatologies.

In a study driven by interest in water resources and agricultural production, it is logical to initially focus on precipitation climatologies. Figure 2 shows the ensemble mean reanalysis precipitation climatology and the individual contributions (as normalised differences). In addition to annual totals, seasonal precipitation is differentiated between a cold season, October to March known regionally as the "rabi", and the monsoon season, April to September referred to as the "kharif". The regional dominance of monsoonal rainfall is striking when comparing the ensemble means of the seasonal contributions to annual total precipitation; although for the Karakoram/Hindu Kush and north-western Central Asian deserts the "rabi" precipitation outweighs monsoonal inputs. In comparing the climatologies of the individual reanalyses, the most prominent differences are located along the southern flank of the Himalayan arc and over the Ganges-Brahmaputra Delta along with uplands along the India-Burma border region. Broadly, JRA-55 is drier than the other reanalyses along the Nepal-Bhutan-China border but much wetter over the Terai, Assam, the lower Ganges basin and the Bay of Bengal. NCEP CFSR has similar characteristics, with the exception of being drier over the Bay of Bengal. ERA-Interim and NASA MERRA show the opposite pattern, with ERA-Interim being much wetter over the Nepal-Bhutan-China border region and NASA MERRA being much drier over the Terai, Assam and Ganges-Brahmaputra Delta.

While adequate moisture inputs from precipitation are prerequisite for both river flows and agricultural production, the role of energy inputs in both the generation of meltwater runoff, from snow and glacial ice, and in driving crop development, through photosynthesis and transpiration, are also critical. Figure 3 shows the ensemble mean climatologies and individual (normalised difference) contributions for Winter (December to February) SW_{net}, Spring (March to May) daily T_{avg} and Summer (June to August) DTR. These temporal aggregates – Winter, Spring and Summer – were selected to identify hydrological regimes – pluvial, nival (snowpack) or glacial – and growing seasons dependent upon thermal conditions. As described in Table 2, all three seasonal values – Winter, Spring, Summer – for each of these variables – T_{avg}, SW_{net} and DTR – were used as input to the classification procedure. Figure 3 shows a single seasonal example of each variable to illustrate the information it contributes. Autumn (September to November) seasonal aggregates were not used as they are very similar to Spring (mirror image) in terms of magnitude and variability and thus not expected to substantially increase information content available to the PCA.

Figure 3 shows that Winter SW_{net} illustrates the influence of seasonal snow-cover via albedo. As expected there is a generally latitudinal gradient, with decreasing SW_{net} moving northward, although the latitudinal gradient is smaller than reductions in net surface absorption in areas with seasonal snow cover. JRA-55 shows generally lower SW_{net} values than the ensemble mean, particularly over south-western Pakistan and the Tibetan plateau. The former difference is likely due to greater reanalysis estimates of cloud radiative effect (CRE) while over Tibet this might be either due to CRE or to higher predicted albedo from greater assumed seasonal snow cover. In contrast JRA-55 shows higher SW_{net} over the Pamir and sections of the high Karakoram and Himalayan arc. This may be either due to assumed lesser seasonal snow-cover (decreased albedo) or estimated clearer sky conditions (decreased CRE). Broadly speaking ERA-Interim and NASA MERRA show the opposite contribution patterns to JRA-55, and hence detailed examination of radiation modulating physical mechanisms, e.g. clear versus overcast conditions, full snow cover versus bare ground, would likely reveal opposing tendencies. Between ERA-Interim and NASA MERRA, the former shows broader and more pronounced decreases in SW_{net} continuously along the Himalayan arc from Pamir through the east of Bhutan to the Sikkim. NCEP CFSR shows a mixed pattern of SW_{net}, agreeing with JRA-55 north of approximately 30°N latitude and more closely corresponding to ERA-Interim and NASA MERRA south of this line.

The ensemble mean climatology of Spring daily T_{avg} displays the expected influence of elevation, with sub-freezing temperatures found roughly above 3000m asl. Like SW_{net}, T_{avg} through the freezing isotherm provides a spatial indication of areas with likely snow cover. More generally, T_{avg} quantifies the available energy to drive melting of snow and ice as well as plant development. Although NASA MERRA is notably warmer than the other three reanalyses over the Indo-Gangetic plains, the largest discrepancies are along Himalayan arc as well as at the transition from the Taklimakan desert to the Tibetan Plateau. JRA-55 and NCEP CFSR are generally colder than the mean along the Himalayan arc but warmer along the northern Tibetan fringe. ERA-Interim is strongly warmer along the Himalayan arc but much cooler over the southern Taklimakan. NASA MERRA has more mixed contributions with relatively limited areas showing substantial departures from the ensemble mean.

Summer DTR is not a direct indicator of energy input to the hydro-climatological system and biosphere. It does, however, provide a measure of the amplitude of energy variation throughout the diurnal cycle as well as providing a proxy for relative humidity (vapour pressure deficit) and cloud cover. Examination of the ensemble mean Summer DTR climatology clearly illustrates the influence of both cloud cover and humidity. Regionally Summer DTR is lowest over the Arabian Sea and Bay of Bengal and highest over the western Central Asian deserts. Suppression of Summer DTR is clearly evident by comparing the ensemble mean Summer DTR in Figure 3 to the ensemble mean monsoonal precipitation accumulations in Figure 2. The influence of diurnal discretisation (sub-daily time-step) on individual reanalysis DTR climatologies is evident in Figure 3. NASA MERRA, with an hourly time-step, has much larger DTR values over land than the ensemble mean, although lower DTR values than the mean over the Arabian Sea and Bay of Bengal. MERRA's hourly time-step allows better representation of the full amplitude of the DTR, while the 6-hour time-steps of the other reanalyses "flatten" or dampen estimated diurnal variations. NCEP CFSR has the lowest DTR values, with particularly small DTR estimates over the Central Asian deserts and Tibetan Plateau. ERA-Interim has broadly, if moderately, lower DTR values than the mean except over the Central Asian deserts as well as the Arabian Sea and Bay of Bengal. JRA-55 is similar to ERA-Interim in DTR estimates albeit spatially more variable and closer to the ensemble mean.

In summary, the substantial differences, illustrated in Figures 2 and 3, in input variable climatologies between the individual reanalyses can be attributed to differences in spatial resolution and sub-diurnal discretisation. Reanalyses will also differ in the data assimilation systems and data analysis and forecasting models they incorporate, an exploration of which is beyond the scope of this study. Spatial resolution will have the most pronounced influence in areas with steep topographic gradients and in interface zones between land and sea. Sub-diurnal time-step influence will be limited to absolute accuracy of DTR. While both spatial resolution and sub-diurnal time-step influence absolute accuracy and hence the direct comparability of a reanalysis to other datasets, its internal coherence, i.e. relative spatial and temporal variability, may still be substantial. This coherence can be tested through the climate classification process. Where good ground-based observations exist and can be translated meaningfully to the grid cell resolution in the reanalyses, bias assessment could be performed. This would provide insight into which dataset more accurately represents regional conditions but would be very challenging and time consuming due to data paucity and inconsistencies. This in fact highlights one of the major benefits of the climate classification procedure: objective delineation of the regional domain should enable optimisation of the use of limited ground data by defining “areas of relevance” within which the magnitude and distribution of bias can be meaningfully summarised.

[2.3] Method for climate classification

The climate classification methodology used in this study directly transfers the method developed by Blenkinsop et al. (2008) for the European FOOTPRINT project albeit with the set of variables described in Section 2.2 rather than those identified for FOOTPRINT (Nolan et al., 2008). Blenkinsop et al. (2008) applied a three-step approach to climate zoning: i) identification of key climatic variables, ii) principal components analysis (PCA) and iii) k-means cluster analysis. The decision to use the PCA and k-means approach, which classifies the spatial domain based on relative differences, rather than to apply a classification based on absolute thresholds, e.g. Köppen-Trewartha (Belda et al., 2014), was made due to the expectation that the spatial aggregation (large grid cells) within the reanalyses would introduce inevitable biases. These biases could be further exacerbated by the formulation of data assimilation and forecasting algorithms adopted by each reanalysis. Thus it seemed more reasonable to apply a relative differentiation rather than an absolute, fixed standard.

As explained by Blenkinsop et al. (2008), PCA is a necessary step in the climate classification process in order to reduce the dimensionality of the input variables which are expected to be substantially correlated as a set. Prior to PCA all input variables were standardised (subtraction of spatial mean and division by spatial standard deviation). Standardisation was performed so that the unit-dependent absolute values of the individual variables would not distort their weighting within the PCA process. PCA was performed using the “mlab” module of matplotlib (Hunter, 2007) executed in a Python environment. Input and output operations of reanalysis data stored as GeoTiffs were handled using the RasterIO Python module (Holderness, 2011).

The results of the PCA for each reanalysis are summarised in Table 3. A decision was made to retain principal components (PCs) which accounted for at least 5% of the total variance in the input dataset. Table 3 indicates that ERA-Interim and NCEP CFSR each had 4 PCs which met this criterion while JRA-55 and NASA MERRA had 5 PCs. Details on the first 3 PCs, which together account for between 81% and 85% of the total variance, for each reanalysis are provided in Table 3, while Figure 4 shows these PCs graphically. The first PC for all four reanalyses was primarily composed of variables related to energy inputs (daily mean temperature, net shortwave radiation), although JRA-55, ERA-Interim and NASA MERRA all had substantial negative contributions from Summer DTR. The first PC accounted for between 36% and 46% of the total variance depending on the reanalysis chosen. As can be seen in Figure 4, the differences between the reanalyses in spatial distribution of PC1 within the domain can be largely accounted for by the respective differences in spatial resolution. Even without allowing for the spatial resolution, differences in the consistency in PC1 between reanalyses is striking.

For the second and third PCs, contributions were very similar between three of the reanalyses (Table 3). For ERA-Interim, NASA MERRA and NCEP CFSR, PC2 was dominated by precipitation inputs from all seasons while negative contributions from Summer energy inputs were also present. In these reanalyses PC3 was dominated by DTR, particularly Winter and Spring. For JRA-55, PC2 was dominated by Winter and Spring DTR with a negative contribution from cold season (“rabi”) precipitation. JRA-55 PC3 was dominated by annual total and monsoonal (“kharif”) precipitation as well as Winter DTR. Despite the differences in composition, i.e. loadings from input variables, spatial variability within the domain for PC2 from JRA-55 is visually very similar to PC2 from the other three reanalyses. In PC2, for JRA-55 the Arabian Sea shares the same sign as the Himalayan arc and Ganges-Brahmaputra Delta while in the other three reanalyses the Arabian Sea has the same sign as the Lower Indus Basin and Central Asian deserts. There are more substantial differences between reanalyses in PC3. In JRA-55 the signs of Central Asian deserts and Tibetan plateau are reversed compared to the patterns found in PC3 in the other three reanalyses. For all reanalyses PC2 accounted for between 19% and 32% of total variance while PC3 accounted for between 16% and 19%. Overall the spatial patterns in Figure 4 are physically plausible, especially PC1 (mean annual temperature/energy input) and PC2 (annual total precipitation) in the three similar reanalyses (excluding JRA-55). Spatial patterns in PC3 (cold season/ “rabi” DTR) are also physically plausible, although visually are less intuitive as diurnal temperature cycles are substantial even in high elevation areas (Karakoram, Himalaya, Tibetan Plateau) in these seasons. They are of lesser amplitude, however, than those experienced currently in the Indo-Gangetic plains and Central Asian deserts.

K-means cluster analysis was also performed using matplotlib (Hunter, 2007) and RasterIO (Holderness, 2011) within a Python environment. As suggested by Blenkinsop et al. (2008), standardised grid cell latitude and longitude were added to the retained principal components as input to the clustering process. Because k-means cluster analysis presupposes the number of distinct (climate) classes rather than determining the number groupings (zones) based on a numerical measure of “likeness” a range of cluster numbers was tested for each reanalysis. The results are presented in the following section, but the our interpretation was that the study domain could be aptly described by eight sub-regional climate zones with increases in cluster numbers leading to sub-divisions of these zones. The issue of spatial discretisation of steep topographic gradients, and hence temperature and precipitation gradients, in the transition zone between the (southern flank of the) Himalayan arc and Indo-Gangetic plains does, however, raise a legitimate caveat to this generalisation.

[3] Results

[3.1] Description of emergent regional climate zones and subdivisions

Figure 5 shows the results of k-means clustering for each reanalysis for eight, twelve and sixteen clusters. Similar sub-divisions of the eight sub-regional climate zones tend to emerge in all the reanalyses as cluster numbers increase although sub-divisions first emerge dependent upon spatial discretisation and climatological differences – illustrated in Figures 2 and 3 – of each reanalysis.

The general characteristics of the eight emergent sub-regional climate zones are described Table 4 along with the fraction of the spatial domain each covers in each reanalysis (for the 8-cluster case). With the exception of the Himalayan arc zone which was not identified by both JRA-55 and NASA-MERRA when the number of clusters was limited to eight, there is substantial agreement not only on the broad geographic locations of the eight zones but on their spatial extent within the domain as well. There is arguably some blurring in the definition of the “Lower Indus Basin” (semi-arid plains), which regionally could be seen as a transitional zone between the “Central Asian deserts” and the “Gangetic Plains” (sub-humid plains), although the latter could itself be seen as a transitional zone between the Lower Indus and the “Ganges-Brahmaputra Delta” (humid plains).

[3.2] Comparison of climatologies of emergent sub-regional climate zones

The spatial mean and ranges (minimum and maximum) have been calculated for the period monthly means of the four input variables from each reanalysis. The annual cycles of precipitation and DTR are shown in Figure 6. The annual cycles of daily mean temperature and net shortwave radiation are shown in Figure 7. Placement of sub-regional zones within these figures are deliberate in their relationship to geographical location and large-scale circulation influences. The most northerly zones are in the upper figure panels and the southerly at the bottom. Zones with greater westerly weather system influence are in the left hand column, while greater monsoonal influence zones are to the right. Results shown in both figures are referred to in the discussion throughout this Section.

[3.2.1] Precipitation climatologies of emergent sub-regional climate zones

Precipitation is a core element in differentiating the eight emergent sub-regional climate zones within the study domain. The Ganges-Brahmaputra Delta (humid plains) has by far the highest precipitation of the sub-regional zones followed by the Gangetic plains (sub-humid plains) and the Himalayan arc. Precipitation in each of these zones is dominated by monsoonal rainfall although the Himalayan arc receives moderate precipitation from westerly weather systems in late winter (February) and Spring. The Karakoram/Hindu Kush zone is the next wettest with dominant inputs from "rabi" westerly weather systems and limited Summer rainfall. The Tibetan Plateau has a similar seasonal distribution of precipitation to the Himalayan arc but with lower monthly totals. The Lower Indus Basin and Central Asian deserts are the driest zones. Spread in spatial means between reanalyses is substantial for all climate zones and appears roughly proportional to precipitation amount, i.e. the largest spread is found in the wettest months and in the wettest zone (Ganges-Brahmaputra Delta).

[3.2.2] DTR climatologies of emergent sub-regional climate zones

As explained in Section 2.2, ensemble spread in DTR climatologies can be substantially attributed to issues of sub-diurnal discretisation. For all climate zones except the Arabian Sea and Bay of Bengal, the reanalysis with an hourly time-step (NASA MERRA) has the largest DTR values. Despite similar sub-diurnal discretisation, NCEP CFSR has consistently lower DTR values across all climate zones than ERA-Interim and JRA-55 which tend to agree closely with one another. Despite this considerable ensemble spread in absolute values, the "shape" of annual DTR cycles within climate zones is consistent between reanalyses, i.e. standardised values are very similar. Zones with substantial monsoonal influence – Ganges-Brahmaputra Delta, Gangetic plains and Himalayan arc – have annual DTR minima in Summer. In contrast, drier and more westerly-dominated sub-regional zones – Central Asian deserts, Tibetan plateau, Karakoram/ Hindu Kush and Lower Indus Basin – have annual DTR minima in Winter, although the Lower Indus has sufficient monsoonal influence for a minor minimum (limited DTR suppression) in Summer. The Arabian Sea and Bay of Bengal have the smallest DTR values both in absolute terms (annual mean) and amplitude of annual cycle.

[3.2.3] Daily mean temperature climatologies of emergent sub-regional climate zones

Based on the PCA results presented in section 2.3, differences in energy inputs account for the largest fraction of variance within the input data. Differences in annual cycles of daily T_{avg} provide clear differences between the emergent sub-regional climate zones. The Arabian Sea and Bay of

Bengal have year-round moderately warm temperatures with minimal spread in both ensemble mean and in spatial spread within individual reanalyses. The Ganges-Brahmaputra delta has similar monthly spatial mean values to the Arabian Sea but with incrementally larger ensemble spread and much greater spatial spread. The spatial spread is attributed to the topographic diversity within the zone, stretching from coastal areas to the front ranges of the Himalaya. The Lower Indus Basin and Gangetic plains have quite similar annual cycles of daily mean temperature. Both have mild cold seasons ("rabi") and hot summers with large spatial spreads in all months. The ensemble spread is incrementally larger in all months for the Lower Indus than for the Gangetic plains. The remaining four zones– Central Asian deserts, Tibetan Plateau, Karakoram/ Hindu Kush and Himalayan arc – are alike in several months of the annual cycle, with mean temperatures below freezing. Ensemble and spatial spreads are greater in the Central Asian deserts and Karakoram/ Hindu Kush than in the Tibetan Plateau, which is consistently the coolest zone. For the Himalayan arc, ERA-Interim and NCEP CFSR agree closely for both the spatial means and the considerable spatial spreads of this zone.

[3.2.4] Net shortwave radiation climatologies of emergent sub-regional climate zones

Net shortwave radiation at the surface is, understandably, the least differentiated of the input variables. Of interest is the varying degrees of SW_{net} suppression in different seasons. In cold months shortwave suppression is due to increased albedo from seasonal snow cover and to a lesser extent to CRE from thick cloud cover. This is evident in the Tibetan Plateau and Karakoram/ Hindu Kush where the annual minima is well below 100 watts/m^2. Sub-100 watts/m^2 annual minima in the Central Asian deserts are more surprising and may in part be due to airborne dust particles. Higher Winter SW_{net} for the Himalayan arc, comparable to the Lower Indus, than the Karakoram/ Hindu Kush may be attributable to the lower latitude and lesser seasonal snow cover of the more easterly mountain range. Summer SW_{net} suppression will be caused by large CRE linked to monsoonal activity. This is particularly visible in the Ganges-Brahmaputra Delta and Gangetic plains and still noticeable in the Himalayan arc and Arabian Sea. The effect is present though barely perceptible in the Lower Indus Basin.

[3.2.5] Commonalities and distinctions in the climatologies of emergent sub-regional climate zones

The layout of Figures 6 and 7 is intended to facilitate comparison of adjacent climate zones. Climate zones are represented within Figures 6 and 7 moving from north to south by moving from top to bottom panels. Given the latitudinal influence on temperature, zones with similar temperature regimes, e.g. Lower Indus Basin and Gangetic plains, are laterally adjacent. In contrast, the dependence of precipitation on atmospheric circulation can be examined by comparing these adjacent panels. Thus the Lower Indus Basin, with limited monsoonal rainfall is found by the clustering process to be distinct from the Gangetic plains. Similarly the Tibetan plateau is distinguished from the Central Asian deserts not only by cooler temperatures but also by greater monsoonal precipitation. The Karakoram/ Hindu Kush and Himalayan arc have similar temperature regimes but the seasonality and magnitude of annual precipitation, driven by the differing circulation influences, clearly separates them. Even without knowledge of land or sea presence, the Ganges-Brahmaputra Delta zone is distinct from the Arabian Sea zone by both precipitation and DTR.

[4] Discussion

[4.1] Insights from climate classifications for water resources and food security in South Asia

The PCA and k-means clustering approach applied to climate classification for the Himalayan arc and adjacent regions, focusing on water resources and food security, has found a consensus among four global meteorological reanalyses to identify eight emergent sub-regional climate zones. These zones are physically plausible and correspond to broadly recognized units of vegetation typology and land surface characteristics in South and Central Asia. Of these eight zones, one is open water (Arabian Sea and Bay of Bengal) while two – Central Asian deserts and Tibetan Plateau – are sparsely populated. The three plains zones – Lower Indus Basin, Gangetic plains and Ganges-Brahmaputra Delta – are densely populated and projected to experience rapid demographic growth in the coming decades (Archer et al., 2010; Immerzeel and Bierkens, 2012). In addition to direct precipitation assessed in the climate classification these plains regions receive river flows from upstream areas: the Karakoram/ Hindu Kush is upstream of the Lower Indus Basin while the Himalayan arc is upstream of the Gangetic plains and Ganges-Brahmaputra Delta. The precipitation climatologies of individual climate zones presented in Figure 6 confirm that the Lower Indus Basin receives substantially less direct precipitation than the other two plains climate zones. In a first-order analysis, irrigated areas in the Lower Indus, shown in Figure 1, are thus much more dependent upon upstream flows than their Gangetic counterparts.

This general assessment does not however take into account the question of intra-annual (inter-seasonal) water transfers, as the annual cycle of Ganges basin tributary river flows will closely follow the annual precipitation cycle. Thus, in the absence of impounding reservoirs or substantial groundwater recharge, only limited water volumes would be available to supplement irrigation in the dry “rabi” season. This study also does not take into account inter-annual variability, as the climate classifications here draw solely upon period means (1980 to 2009). A further limitation of this assessment is that at the “parcel scale” of rainfed agriculture the convective precipitation in monsoonal weather systems has very large spatial variability (Khan et al., 2014). Thus while farmers in the irrigated Lower Indus Basin rely upon upstream flows for the bulk of crop moisture requirements, farmers in the Gangetic plains may find supplementary irrigation critical to compensate spatially and temporally acute precipitation deficits and ensure crop yields.

Looking forward, climate classifications of the type applied in this study help to frame the assessment of the impact of changing climate conditions on future water resources, crop production and food security. By understanding the roles of sub-regional climate zones as water resource supply (headwaters) and demand (irrigated plains) areas, the net result of changes in water availability (precipitation change) and potential evapotranspiration (air temperature, shortwave radiation and relative humidity change) can be more skilfully evaluated. Changes, calculated between time-slices of dynamically-downscaled climate model simulations, in both the spatial extent and climatological statistics of water resource supply and demand zones in and of themselves provide information on the trajectory of water availability, i.e. unit yield or deficit multiplied by surface area. Additionally, delineation of sub-regional climate zones provides an objective basis for definition of study boundaries of more sophisticated nested downscaling investigations. Accurate delineation is important when computational requirements are high, for example when high-resolution sensitivity experiments are required to constrain the uncertainties in future supply and demand scenarios.

[4.2] Utility of climate classification for assessment of gridded datasets

The ensemble reanalysis input climatologies and normalised difference contributions shown in Figures 2 and 3 illustrate the initial steps in comparative assessment of gridded data sets for bias characterisation and validation. Further logical steps would draw upon the climate zones derived through the PCA and k-means clustering approach to sub-divide the spatial domain in order to focus and organise the use of limited in-situ data (ground-based, point observations) to characterise sub-regional dataset performance. The use of in-situ data to provide “ground truthing” and related large scale datasets to local conditions will remain crucial for the foreseeable future because gridded datasets of a global nature, be they reanalyses, spatially interpolated from local observations or derived from satellite imagery will inevitably have intrinsic biases. These biases are a function of spatial and temporal resolution of the source observations as well as the physical nature of those observations. In-situ data, be they from national monitoring networks or international databases such as the Global Historical Climatology Network (Lawrimore et al., 2011), could be grouped by the derived climate zones and in this way structure the analysis of statistics of “grid-cell versus station” biases. In this way individual gridded datasets could be assessed to determine in which sub-regional climate zones they perform well or poorly. This approach also permits comparative evaluation of different gridded datasets to determine which most accurately reproduces the climatology of a given climate zone.

This proposed methodology for bias assessment is dependent, however, upon the availability of station data which are representative of climatic conditions in absolute terms at the grid-scale level. This constraint could be prohibitive for mountainous areas, such as the Karakoram/ Hindu Kush, where meteorological stations are often located in valley bottoms, substantially below the mean elevations of overlying data source grid cells. One such example is the Upper Indus Basin (Gilgit-Baltistan administrative district of Pakistan) where Archer (2003, 2004) and Archer and Fowler (2004, 2008) found climate observations at manned meteorological stations of the Pakistan Meteorological Department located in valley settlements to correlate strongly with variability in hydrological conditions, although runoff volume fluctuations did not equate directly to precipitation anomalies. Thus in mountainous or other highly spatially variable domains “transfer functions” (scaling relationships) representing climate parameter variation with topography may still be necessary to compare in-situ point observations to grid cell spatial means in absolute terms.

These challenges for relating point-based observations to gridded data in fact point toward the utility of inter-comparison of spatial datasets. The climate classification approach provides a supplementary dimension in which to compare gridded datasets. To illustrate this, the sub-regional climate zones delineated from the four reanalyses could be considered as reference or benchmark values for evaluation of climate model control period outputs. On-going work is exploring the application of the climate classification approach to time-slices within the Met Office Hadley Centre seventeen-member perturbed physics ensemble of 130-year transient future climate simulations (Collins et al., 2011) dynamically downscaled to 0.22 decimal degrees for the South Asia domain (Bhaskaran et al., 2012). Climate classifications, using eight clusters, for the initial 30-years (1970 to 1999) of the simulation, considered as the “control climate”, are shown for each of the ensemble members in Figure 8. Visual comparison of Figure 8 to Figure 5 confirms that the broad patterns of the sub-regional climate zones found by the reanalyses are replicated in the control climate time-slice of the climate model ensemble. There are noteworthy differences, particularly over the Ganges-Brahmaputra Delta, but the overall sub-regional differences are unmistakeable. Table 5 provides the distribution of the spatial domain among the sub-regional climate zones for each climate model ensemble member. The ensemble mean and standard deviation are also given in Table 5. These values are compared, in Table 6, to the equivalent values from the reanalyses (from Table 4). The largest differences in fractional areas stem from an eastern Himalayan climate zone in the model ensemble amalgamating area allocated to the Ganges-Brahmaputra in the reanalyses as well as sections assigned to the Tibetan Plateau in the reanalyses being assigned to the Karakoram Hindu Kush in the model ensemble. Future work will investigate differences in climatology between reanalysis zones (as presented in Section 3.2 and Figures 6 and 7) and the model ensemble zones. This analysis will then be extended to compare climate classifications between time-slices of the model ensemble.

In summary, the climate classification approach presented here has substantial potential both for use in assessment of water resources and food security issues as well as for the characterisation of skill and bias of gridded datasets for reproducing sub-regional climatologies. This relative, or internal-difference, classification approach was preferred over a methodology based on fixed, absolute thresholds due to the nature of the gridded datasets whose spatial discretisation on likely intrinsic biases would distort the results of an absolutist method. The natural resource assessment application of this approach is timely as increasing pressures on water resources and cropland appear inevitable in South Asia for the medium term due to demographic trends and evolving consumption patterns. The growing availability of gridded datasets increases the likelihood of their use to address resource management and climatic sensitivity issues. In order to use these datasets skilfully it is necessary to first rigorously characterise their performance and biases. Thus the climate classification approach presented here is doubly timely as it provides a framework to organise use of in-situ observations to differentiate gridded dataset performance at the sub-regional level and to carry out inter-comparison of gridded dataset performance for these sub-regions.

[5] Conclusions

A three-step approach was used to derive climate classifications for the Himalayan arc and adjacent plains from climate inputs from four global meteorological reanalyses covering the recent historical record (1980 to 2009). Input variables were selected for this process with a focus on climatic drivers of water resources and agricultural production. Knowledge of the climatic factors governing behaviour of hydrological regimes with substantial contributions from seasonal snowpack and glaciers as well as controlling crop growth led to selection of precipitation amount, daily mean temperature, net shortwave radiation at the surface and DTR as input variables. Three seasonal aggregations were chosen for each input variable. Annual, “rabi” (October to March) and or “kharif” (April to September) totals were used for precipitation to differentiate the influences of westerly mid-latitude and monsoonal sub-tropical weather systems. For the remaining variables temporal aggregates for Winter (December to February), Spring (March to May) and Summer (June to August) were selected to identify hydrological regimes – pluvial, nival (snowpack) or glacial – and growing seasons dependent upon thermal conditions.

Principal Components Analysis (PCA) was applied to the spatially standardised temporal aggregates of the input variables. Comparison of PCA results from the four reanalyses show that in all cases the first principal component was dominated by energy inputs while the second and third were dominated by precipitation and DTR. Principal components accounting for a minimum of 5% of total input variance, supplemented with standardised latitude and longitude, were used as inputs to a k-means cluster analysis. Progressive increases in cluster numbers were tested for each reanalysis in order to assess the evolution of emergent climate zones. Results of the k-means analysis were interpreted to show that the study domain could be adequately described by eight sub-regional climate classifications while further increases in cluster numbers resulted in sub-divisions of these macro-zones. Spatial statistics for each sub-regional climate zone from the ensemble of reanalyses revealed consistent, distinct climatologies in the annual cycles of the input variables.

The capacity of the climate classifications to provide insight into water resources and food security issues at a regional scale were discussed. This capacity is linked to the objective delineation of water resource supply and demand zones. Analysis of changes in both the spatial and climatic characteristics of the zones over time provides a framework for evaluation of water availability for crop production. The climate classifications also support evaluation of gridded datasets themselves. The climate zones provide an objective method for grouping available ground-based observations to quantify and summarise gridded dataset bias. They also serve as a metric with which to compare climatologies of gridded datasets. This was illustrated by comparing the climate classifications of the ensemble of reanalyses to the "control period" of a dynamically downscaled perturbed physics climate model ensemble. Strong commonalities between the benchmark (reanalysis) and predictive (RCM) datasets were evident while limited divergences were clearly identified. Future work will extend the methodology here to evaluate the regional water resources and food security implications of changes projected by available RCM experiments covering South Asia and the Himalayan arc.

[X] Acknowledgements

This study was made possible by financial support from the Leverhulme Trust via a Philip Leverhulme Prize (2011) awarded to Prof Fowler. Local meteorological data was acquired from the Pakistan Meteorological Department (PMD) with support from the Global Change Impacts Studies Centre (GCISC). Additional financial support during the developmental stages of this work was provided by the British Council (PMI2 and INSPIRE grants), the UK Natural Environment

Research Council (NERC) as a Postdoctoral Fellowship award NE/D009588/1 (2006–2010) to Prof Fowler, and a US National Science Foundation (NSF) Graduate Research Fellowship award to Dr Forsythe (2006–2010).

References

Archer, D. R., 2003. Contrasting hydrological regimes in the Indus Basin. J. Hydrol., 274 (1-4), 198–210. http://dx.doi.org/10.1016/S0022-1694(02)00414-6

Archer, D. R., 2004. Hydrological implications of spatial and altitudinal variation in temperature in the Upper Indus Basin. Nord. Hydrol, 35 (3), 209–222.

Archer, D.R, Forsythe, N., Fowler, H.J. and Shah, S.M., 2010. Sustainability of water resources management in the Indus Basin under changing climatic and socio economic conditions. Hydrol. Earth Syst. Sci. 14, 1669–1680. http://dx.doi.org/10.5194/hess-14-1669-2010

Archer, D. R. and Fowler, H. J., 2004. Spatial and temporal variations in precipitation in the Upper Indus Basin, global teleconnections and hydrological implications. Hydrol. Earth Syst. Sci., 8 (1), 47–61. http://dx.doi.org/10.5194/hess-8-47-2004

Archer, D. R. and Fowler, H. J., 2008. Using meteorological data to forecast seasonal runoff on the River Jhelum, Pakistan, J. Hydrol., 361 (1-2), 10–23. http://dx.doi.org/10.1016/j.jhydrol.2008.07.017

Baigorria, G.A., Jones, J.W., Shin, D.-W., Mishra, A. and O'Brien, J.J., 2007. Assessing uncertainties in crop model simulations using daily bias-corrected Regional Circulation Model outputs. Climate Research, 34 (3), 211-222. http://dx.doi.org/10.3354/cr00703

Belda, M., Holtanová, E., Halenka, T. and Kalvová, J., 2014. Climate classification revisited: From Köppen to Trewartha. Climate Research, 59 (1), 1-13. http://dx.doi.org/10.3354/cr01204

Bhaskaran, B., Ramachandran, A., Jones, R. and Moufouma-Okia, W., 2012. Regional climate model applications on sub-regional scales over the Indian monsoon region: The role of domain size on downscaling uncertainty. Journal of Geophysical Research: Atmospheres, 117 (10), art. no. D10113. http://dx.doi.org/10.1029/2012JD017956

Blenkinsop, S., Fowler, H.J., Dubus, I.G., Nolan, B.T., Hollis and J.M., 2008. Developing climatic scenarios for pesticide fate modelling in Europe. Environmental Pollution, 154 (2), 219-231. http://dx.doi.org/10.1016/j.envpol.2007.10.021

Chen, D. and Chen, H.W., 2013. Using the Köppen classification to quantify climate variation and change: An example for 1901-2010. Environmental Development, 6 (1), 69-79. http://dx.doi.org/10.1016/j.envdev.2013.03.007

Collins, M., Booth, B.B.B., Bhaskaran, B., Harris, G.R., Murphy, J.M., Sexton, D.M.H. and Webb, M.J., 2011. Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles. Climate Dynamics, 36 (9-10), 1737-1766. http://dx.doi.org/10.1007/s00382-010-0808-0 de Fraiture, C., Wichelns, D., 2010. Satisfying future water demands for agriculture. Agricultural Water Management, 97 (4), 502-511. http://dx.doi.org/10.1016/j.agwat.2009.08.008

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Holm, E.V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J.-N. and Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137 (656), 553–597. http://dx.doi.org/10.1002/qj.828

Easterling, D.R., Horton, B., Jones, P.D., Peterson, T.C., Karl, T.R., Parker, D.E., Salinger, M.J., Razuvayev, V., Plummer, N., Jamason, P. and Folland, C.K., 1997. Maximum and minimum temperature trends for the globe. Science 277 (5324), 364–367. http://dx.doi.org/10.1126/science.277.5324.364

Ebita, A., Kobayashi, S., Ota, Y., Moriya, M., Kumabe, R., Onogi, K., Harada, Y., Yasui, S., Miyaoka, K., Takahashi, K., Kamahori, H., Kobayashi, C., Endo, H., Soma, M., Oikawa, Y. and Ishimizu, T.,, 2011. The Japanese 55-year Reanalysis “JRA-55”: An Interim Report. Scientific Online Letters on the Atmosphere, 7, 149-152. http://dx.doi.org/10.2151/sola.2011-038

Elguindi, N., Grundstein, A., Bernardes, S., Turuncoglu, U. and Feddema, J., 2013. Assessment of CMIP5 global model simulations and climate change projections for the 21st century using a modified Thornthwaite climate classification. Climatic Change, pp. 1-16. http://dx.doi.org/10.1007/s10584-013-1020-0

Holderness, T. 2011. Python Spatial Image Processing (PyRaster). https://github.com/talltom/PyRaster

Hunter, J. D., 2007. Matplotlib: A 2D graphics environment. Computing In Science & Engineering 9, 90-95. http://dx.doi.org/10.1109/MCSE.2007.55

Immerzeel, W.W., Pelicciotti, F., and Shrestha, A. B. 2012. Glaciers as a Proxy to Quantify the Spatial Distribution of Precipitation in the Hunza Basin. Mountain Research and Development, 32 (1), 30-38. http://dx.doi.org/10.1659/MRD-JOURNAL-D-11-00097.1

Immerzeel, W.W., Van Beek, L.P.H. and Bierkens, M.F.P., 2010. Climate change will affect the Asian water towers. Science, 328 (5984), 1382–1385. http://dx.doi.org/10.1126/science.1183188

Immerzeel, W.W. and Bierkens, M.F.P., 2012. Asia's water balance. Nature Geoscience, 5 (12), 841-842. http://dx.doi.org/10.1038/ngeo1643

Kar, G., Kumar, A., Sahoo, N., Mohapatra, S., 2014. Radiation utilization efficiency, latent heat flux, and crop growth simulation in irrigated rice during post-flood period in east coast of India. Paddy and Water Environment, 12 (2), 285-297. http://dx.doi.org/10.1007/s10333-013-0381-3

Khan, S.I., Hong, Y., Gourley, J.J., Khan Khattak, M.U., Yong, B. and Vergara, H.J., 2014. Evaluation of three high-resolution satellite precipitation estimates: Potential for monsoon monitoring over Pakistan. Advances in Space Research, 54 (4), Pages 670-684. http://dx.doi.org/10.1016/j.asr.2014.04.017

Lawrimore, J.H., Menne, M.J., Gleason, B.E., Williams, C.N., Wuertz, D.B., Vose, R.S. and J. Rennie, J., 2011. An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3, J. Geophys. Res., 116, D19121. http://dx.doi.org/10.1029/2011JD016187.

Mahlstein, I., Daniel, J.S. and Solomon, S., 2013. Pace of shifts in climate regions increases with global temperature. Nature Climate Change, 3 (8), 739-743. http://dx.doi.org/10.1038/nclimate1876

Nolan, B.T., Dubus, I.G., Surdyk, N., Fowler, H.J., Burton, A., Hollis, J.M., Reichenberger, S. and Jarvis, N.J., 2008. Identification of key climatic factors regulating the transport of pesticides in leaching and to tile drains. Pest Management Science, 64 (9), 933-944. http://dx.doi.org/10.1002/ps.1587

Ragettli, S., Pellicciotti, F., Bordoy, R. and Immerzeel, W.W., 2013. Sources of uncertainty in modeling the glaciohydrological response of a Karakoram watershed to climate change. Water Resources Research, 49 (9), 6048-6066. http://dx.doi.org/10.1002/wrcr.20450

Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster, R.D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C.R., Reichle, R., Robertson, F.R., Ruddick, A.G., Sienkiewicz, M. and Woollen, J., 2011. MERRA: NASA's modern-era retrospective analysis for research and applications. Journal of Climate, 24 (14), 3624-3648. http://dx.doi.org/10.1175/JCLI-D-11-00015.1

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M.H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Van Den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R.W., Rutledge, G. and Goldberg, M., 2011. The NCEP climate forecast system reanalysis. Bulletin of the American Meteorological Society, 91 (8), pp. 1015-1057. http://dx.doi.org/10.1175/2010BAMS3001.1

Siebert, S., Döll, P., Feick, S., Hoogeveen, J. and Frenken, K., 2007. The digital global map of irrigation areas (version 4.5). Food and Agriculture Organization of the United Nations, Rome, Italy. http://www.fao.org/nr/water/aquastat/irrigationmap/index10.stm

Uppala, S.M., P.W. Kallberg, A.J. Simmons, U. Andrae, V. Da Costa Bechtold, M. Fiorino, J.K. Gibson, J. Haseler, A. Hernandez, G. A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R.P. Allan, E. Andersson, K. Arpe, M. A. Balmaseda, A.C. M. Beljaars, L. Vande Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Holm, B.J. Hoskins, L. Isaksen, P.A.E.M. Janssen, R. Jenne, A.P.McNally, J.-F. Mahfouf, J.-J. Morcrette, N. A. Rayner, R.W. Saunders, P. Simon, A. Sterl, K.E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo J. Woollen. 2005. The ERA-40 re-analysis. Quarterly Journal of the Royal Meteorological Society 131 (612) 2961-3012. http://dx.doi.org/10.1256/qj.04.176

Zhang, X. and Yan, X., 2014. Spatiotemporal change in geographical distribution of global climate types in the context of climate warming. Climate Dynamics, 43 (3-4), 595-605. http://dx.doi.org/10.1007/s00382-013-2019-y

Table 1. Reanalysis datasets utilised for comparative climate classification.

Reanalysis	Producer	Time period covered	Spatial resolution (degrees)	Diurnal discretisation
JRA-55	JRA	1958 to (near) present	1.25x1.25	6-hour synoptic forecast/analysis periods
ERA-Interim	ECMWF	1979 to (near) present	0.75x0.75	6-hour synoptic forecast/analysis periods
CFSR	NCEP	1979 to 2009 (later extended)	0.50x0.50	6-hour synoptic forecast/analysis periods
MERRA	NASA	1979 to (near) present	0.67x0.50	hourly

Table 2. Variables used for Himalayan region climate classification.

Variable	Season	Physical importance
Precipitation	Annual Total	Humid vs arid climates
	ONDJFM (“rabi”)	Westerly (extra-tropical) weather system climate influence
	AMJJAS (“kharif”)	Monsoonal weather system climate influence
T_{avg} daily mean near surface air temperature	DJF	Indicator of precipitation state (solid versus liquid) and available energy to drive hydrological processes (meltwater generation) and crop growth (transpiration); as such indicator of hydrological regime (pluvial, nival or glacial)
	MAM	
	JJA	
DTR diurnal temperature range	DJF	(inverse) Indicator of moisture conditions, i.e. relative humidity and cloud cover, as both suppress DTR; as such proxy for cloud cover further informs regarding circulation influences
	MAM	
	JJA	
SW_{net} at surface net downward shortwave radiation at the surface	DJF	Indicator of land surface state (snow covered or bare) and available energy to drive hydrological processes (meltwater generation) and crop growth (transpiration) ; as such indicator of hydrological regime (pluvial, nival or glacial)
	MAM	
	JJA	

Table 3. Comparison of results of Principal Components Analysis.

Gridded data source		**PC1**	**PC2**	**PC3**
JRA-55	Explained Variance	**0.459**	**0.194**	**0.162**
5 PCs > 0.05	Loading	JJA DTR -0.359 DJF T_{avg} 0.380 DJF SW_{net} 0.384	ONDJFM Precip -0.440 DJF DTR 0.408 MAM DTR 0.509	AnnTot Precip -0.419 AMJJAS PrecipTot -0.416 DJF DTR -0.461
ERA-Interim	Explained Variance	**0.364**	**0.317**	**0.167**
4 PCs > 0.05	Loading	JJA DTR -0.353 DJF T_{avg} 0.443 MAM T_{avg} 0.404 DJF SW_{net} 0.402	AnnTot Precip 0.460 AMJJAS Precip 0.440 ONDJFM Precip 0.407 MAM SW_{net} -0.353 JJA SW_{net} -0.371	DJF DTR 0.622 MAM DTR 0.621
NASA MERRA	Explained Variance	**0.416**	**0.214**	**0.185**
5 PCs > 0.05	Loading	JJA DTR -0.378 DJF T_{avg} 0.404 MAM T_{avg} 0.375 DJF SW_{net} 0.388	AnnTot Precip 0.491 AMMJAS Precip 0.439 ONDJFM Precip 0.479 JJA T_{avg} -0.395	DJF DTR -0.631 MAM DTR -0.635
NCEP CFSR	Explained Variance	**0.377**	**0.275**	**0.181**
5 PCs > 0.05	Loading	DJF T_{avg} 0.451 MAM T_{avg} 0.429 JJA T_{avg} 0.363 DJF SW_{net} 0.424 MAM SW_{net} 0.382	AnnTot Precip 0.459 AMJJAS Precip 0.440 ONDJFM Precip 0.367 JJA SW_{net} -0.429	DJF DTR -0.478 MAM DTR -0.645 JJA DTR -0.462

nb: Rows labelled “Explained Variance” indicate fraction of total input variance accounted for by the Principal Component (PC). Rows labelled “loading” indicate input variables whose (coefficient) contribution to the PC is > 0.35. Loading coefficients are shown with their signs to differentiate between variables with opposing contributions.

Table 4. Description of primary Himalayan region climate zones (8 clusters).

Regional climate zone name/area	Climate type	Characteristics	Fraction of domain covered			
			JRA-55	ERA-Interim	NASA MERRA	NCEP CFSR
Arabian Sea & Bay of Bengal	Sub-tropical ocean	Year-round warm temperatures; Minimal DTR; limited monsoonal precipitation	0.069	0.077	0.066	0.080
Central Asian deserts	Mid-latitude desert	Cold winter; hot summer; Minimal annual precipitation	0.199*	0.150	0.168	0.101
Tibetan Plateau	High elevation desert	Cold winter; mild summer; limited monsoonal precipitation	0.229	0.207	0.266*	0.227
Himalayan arc	Sub-tropical high mountains	Cold winter; mild summer; Substantial monsoonal precipitation weather	**	0.061	**	0.039
Karakoram/ Hindu Kush	Mid-latitude high mountains	Cold winter; mild summer; Substantial precipitation from westerly weather systems (Winter and Spring)	0.058	0.064	0.050	0.064
Lower Indus Basin	Semi-arid plains	Mild winter (cold season); hot summer; limited monsoonal precipitation	0.133	0.152	0.179	0.194
Gangetic Plains	Sub-humid plains	Mild winter (cold season); hot summer; substantial monsoonal precipitation	0.217	0.192	0.163	0.222
Ganges-Brahmaputra Delta	Humid plains	Mild winter (cold season); warm summer; intense monsoonal precipitation	0.090	0.093	0.104	0.069

*: Combination of two climate zones in this reanalysis; **: Not identified by this reanalysis

Table 5. Variability of primary Himalayan region climate zones (8 clusters) area in the Hadley Centre downscaled perturbed physics ensemble , Regionally Quantify Uncertainty in Model Predictions (RQUMP), ~~RQUMP~~ for South Asia.

Ensemble member	**Indian Ocean**	**Central Asian Desert**	**Gangetic Plains**	**Lower Indus Basin**	**Karakoram Hindu Kush**	**Himalayan arc**	**Ganges Brahmaputra Delta**	**Tibetan Plateau**
rqump00	0.062	0.152	0.236	0.169	0.113	0.092	0	0.171
rqump01	0.075	0.15	0.227	0.184	0.104	0.083	0	0.173
rqump02	0.074	0.15	0.251	0.160	0.102	0.080	0	0.180
rqump03	0.074	0.153	0.231	0.173	0.114	0.091	0	0.160
rqump04	0.071	0.145	0.193	0.168	0.135	0.026	0.083	0.175
rqump05	0.064	0.149	0.179	0.157	0.127	0.039	0.093	0.187
rqump06	0.061	0.154	0.216	0.167	0.131	0.076	0	0.192
rqump07	0.068	0.15	0.196	0.154	0.126	0.027	0.086	0.190
rqump08	0.062	0.156	0.209	0.153	0.131	0.098	0	0.188
rqump09	0.062	0.168	0.208	0.178	0.120	0.092	0	0.169
rqump10	0.075	0.270	0.267	0	0.130	0.121	0	0.134
rqump11	0.061	0.152	0.202	0.171	0.136	0.092	0	0.183
rqump12	0.062	0.238	0.175	0.115	0	0.128	0	0.280
rqump13	0.091	0.261	0.300	0	0.171	0.035	0.138	0
rqump14	0.063	0.264	0.263	0	0.100	0.099	0	0.209
rqump15	0.062	0.148	0.202	0.160	0.132	0.025	0.085	0.183
rqump16	0.069	0.240	0.190	0.115	0	0.101	0	0.282
mean	**0.068**	**0.182**	**0.220**	**0.130**	**0.110**	**0.076**	**0.028**	**0.179**
standard deviation	0.008	0.048	0.034	0.065	0.044	0.033	0.047	0.059

Table 6. Comparison of RQUMP perturbed physics ensemble climate model sub-regional climate zone distributions to those from the reanalysis ensemble.

Statistic		**Indian Ocean**	**Central Asian Desert**	**Gangetic Plains**	**Lower Indus Basin**	**Karakoram Hindu Kush**	**Himalayan arc**	**Ganges Brahamaputra Delta**	**Tibetan Plateau**
Ensemble means	Climate model	0.068	0.182	0.220	0.130	0.110	0.076	0.028	0.179
	Reanalyses	0.073	0.154	0.198	0.164	0.059	0.050	0.089	0.232
	Difference	**-0.005**	**0.028**	**0.022**	**-0.034**	**0.051**	**0.026**	**-0.061**	**-0.053**
Ensemble standard deviations	Climate model	0.008	0.048	0.034	0.065	0.044	0.033	0.047	0.059
	Reanalyses	0.006	0.041	0.027	0.027	0.006	0.015	0.014	0.024
	Difference	0.002	0.007	0.007	0.038	0.038	0.018	0.033	0.035

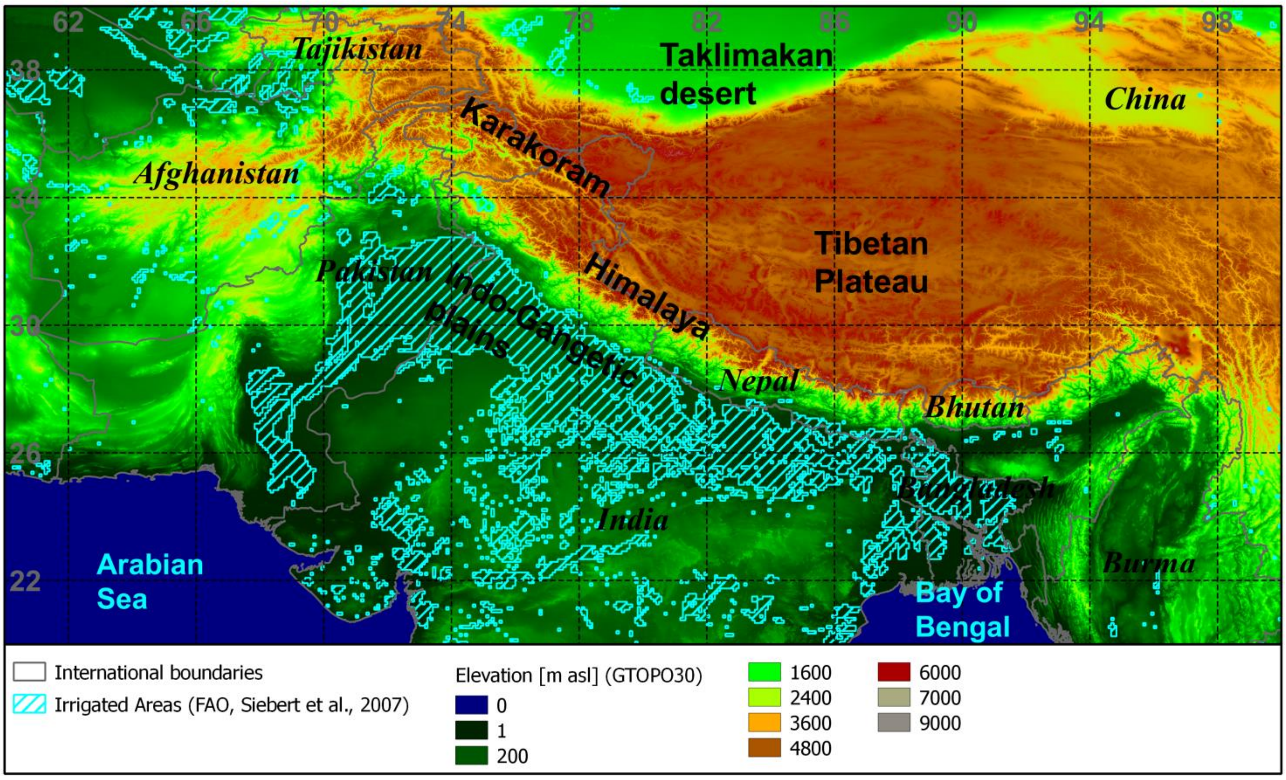

Figure 1. Geographic context of the – Himalayan arc and adjacent plains – study area including elevation and areas with > 33% under irrigation (hashed). Data sources include the United Nations Food and Agriculture Organisation (FAO) and the United States Geological Survey Global 30 Arc-Second Digital Elevation Model (GTOPO30).

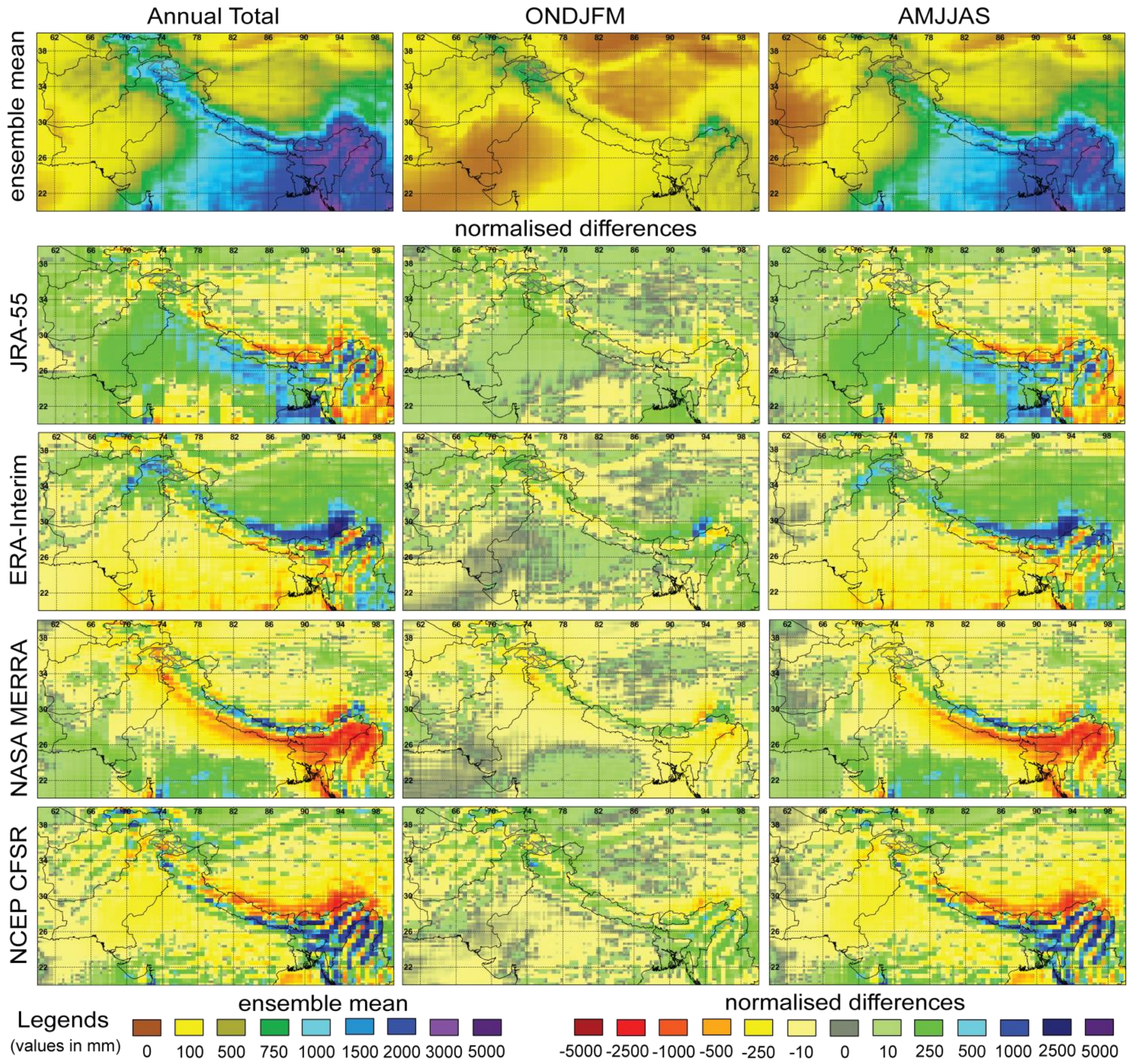

Figure 2. Ensemble precipitation climatology and normalised comparison of individual contributions from reanalyses used in this study. ONDJFM is the abbreviation for the period from October to March, referred to regionally as “Rabi.” AMMJJAS is the abbreviation for the period from April to September, referred to regionally as “Kharif.”

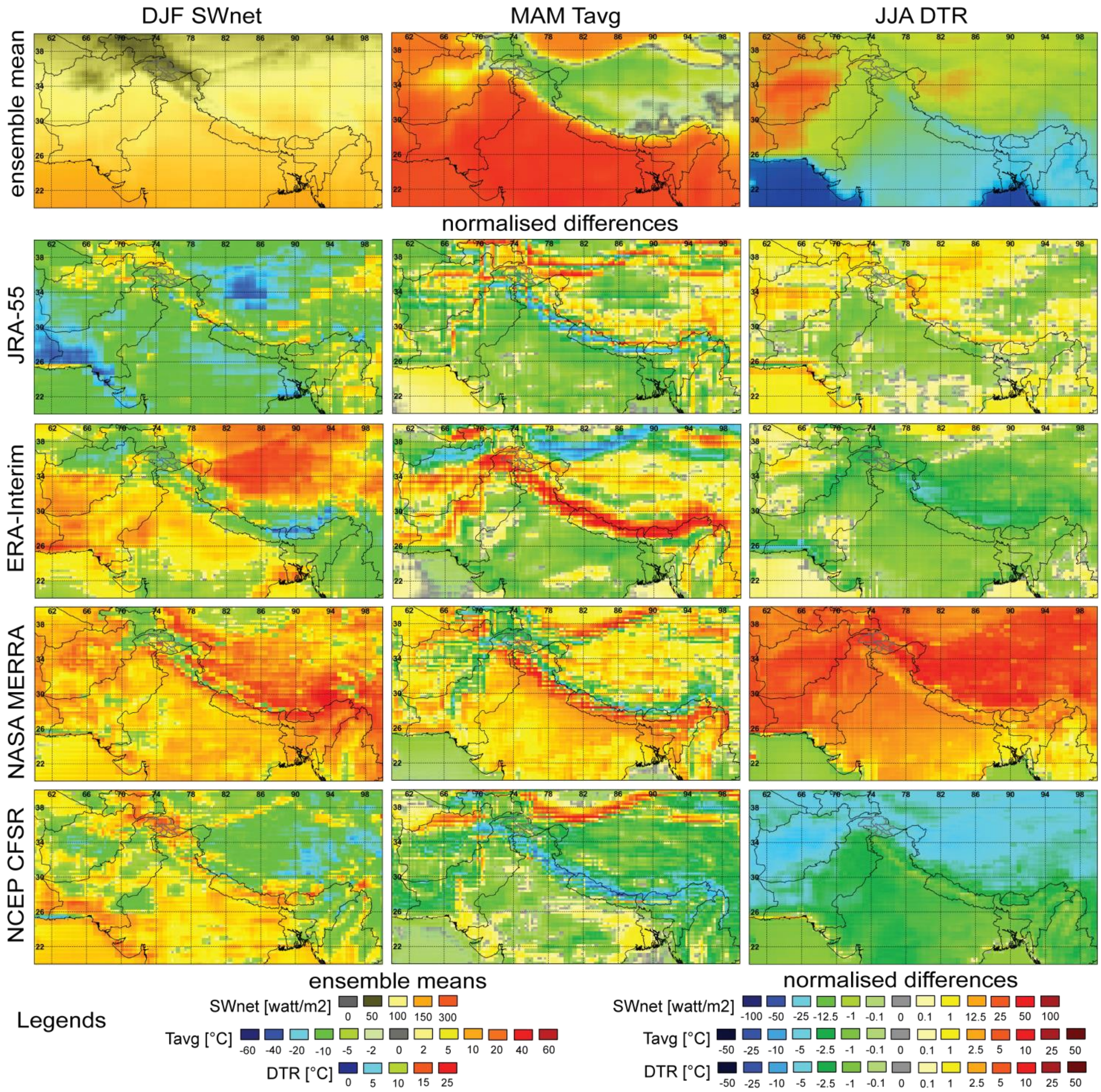

Figure 3. Ensemble energy input (temperature and radiation) climatology and normalised comparison of individual contributions from reanalyses used in this study. SW_{net} is net downward shortwave radiation at the surface. T_{avg} is daily mean near surface air temperature. DTR is diurnal temperature range. DJF is the (Winter) period December through February. MAM is the (Spring) period March through May. JJA is the (Summer) period June through August.

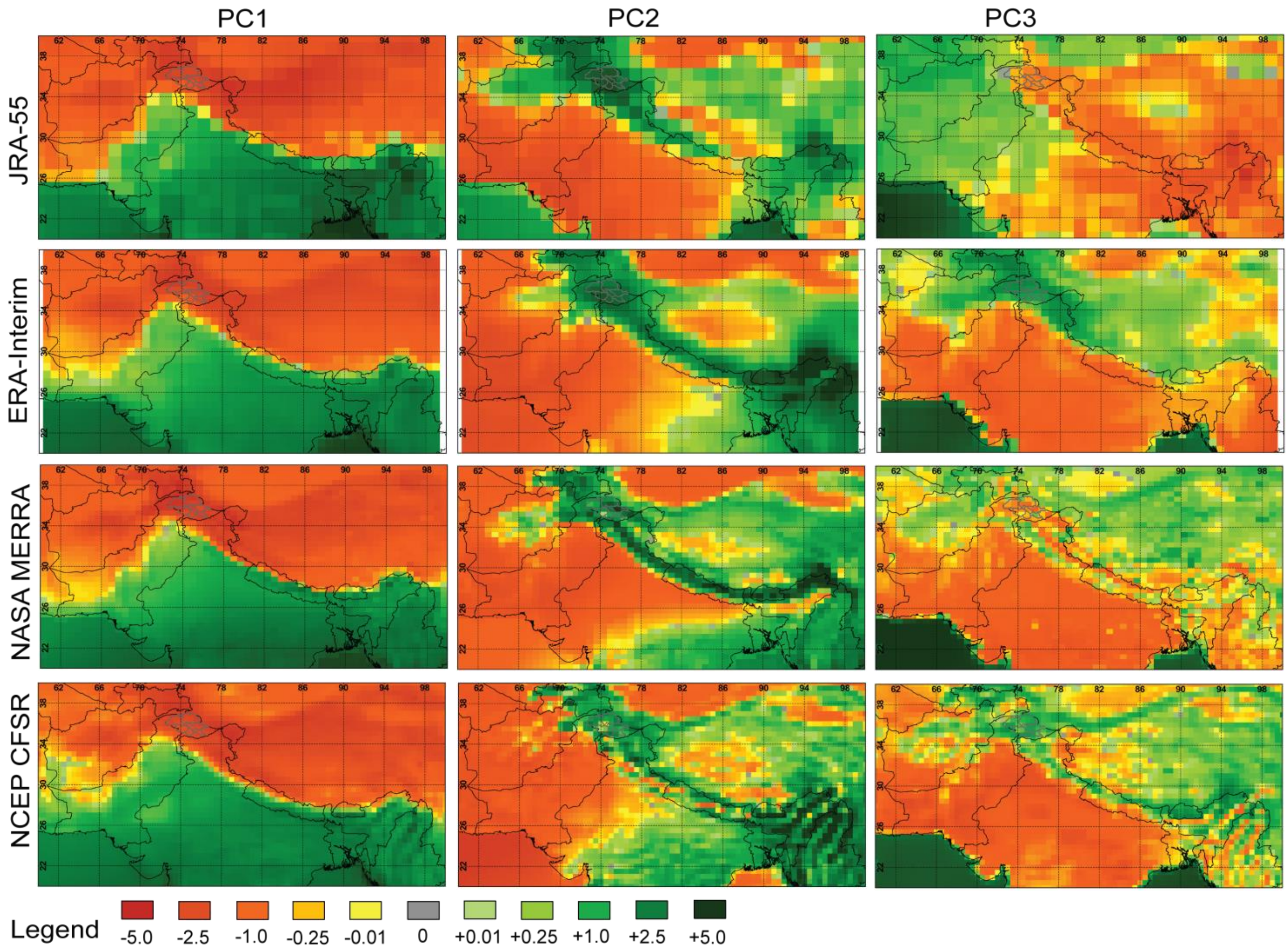

Figure 4. Comparison of the first three principal components (PCs) from each of the reanalyses used in this study. PCs are calculated from the Principal Component Analysis (PCA) input standardised variables using the PCA output weighting factors. PCs are thus dimensionless and values are expressed in standard deviations.

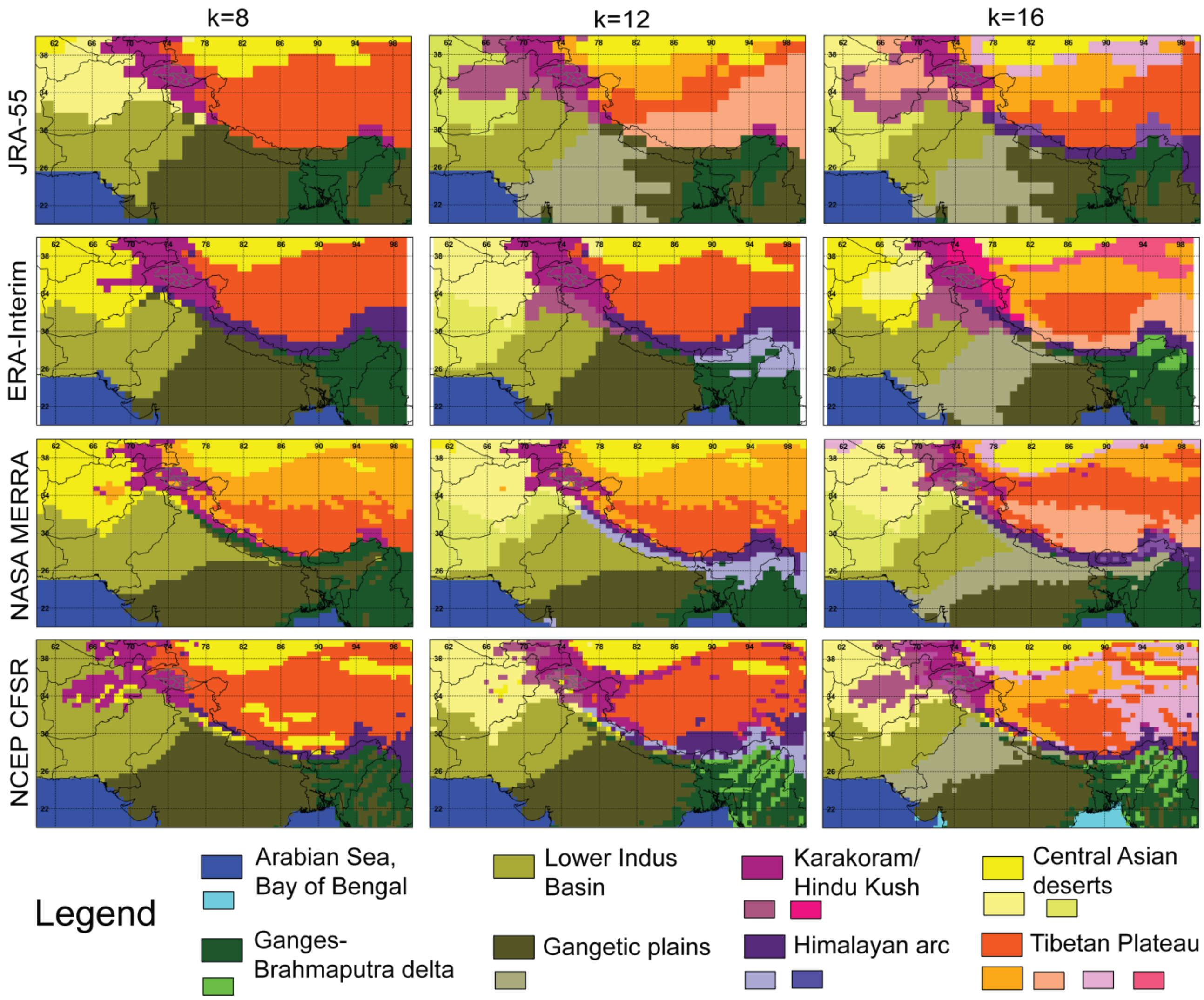

Figure 5. Comparison of climate classifications resulting from the use of 8, 12 and 16 clusters (k) on principal components from the individual reanalyses. Large units in the legend refer to zones for the k=8 case.

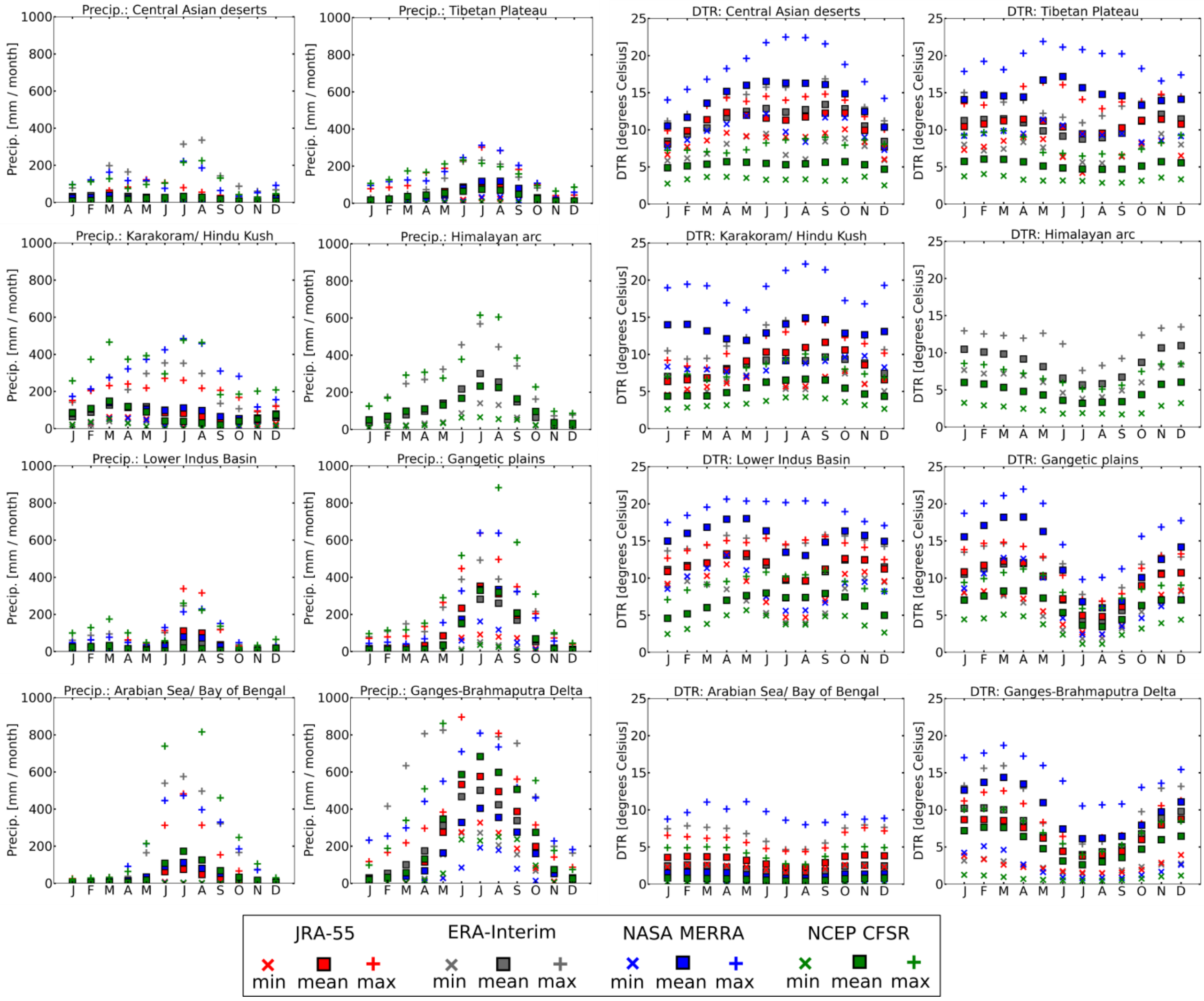

Figure 6. Ensemble spatial statistics for annual cycles of precipitation (left) and DTR (right) by climate zone (8 clusters). DTR is diurnal temperature range.

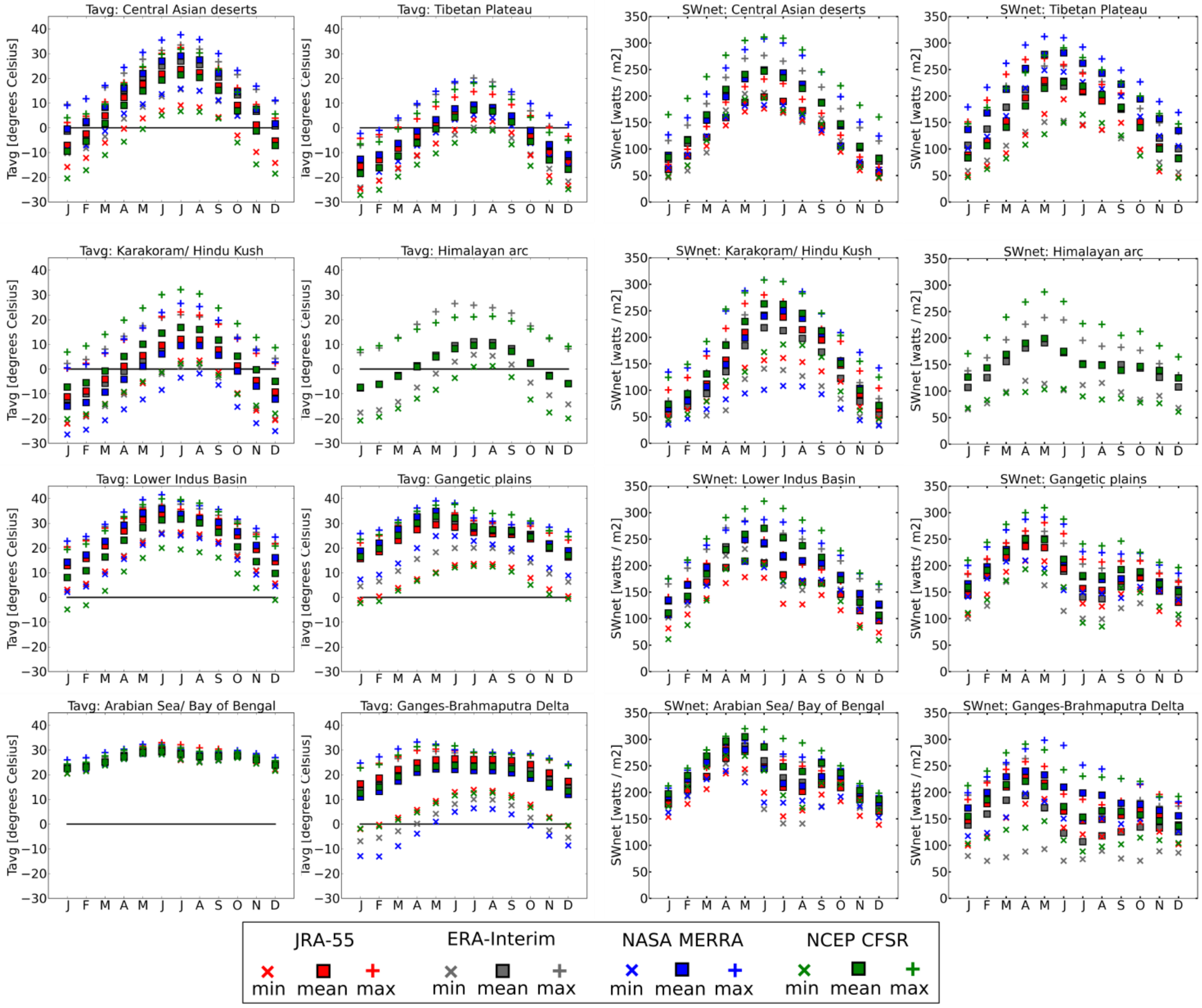

Figure 7. Ensemble spatial statistics for annual cycles of T_{avg} and SW_{net} by climate zone (8 clusters). SW_{net} is net downward shortwave radiation at the surface. T_{avg} is daily mean near surface air temperature.

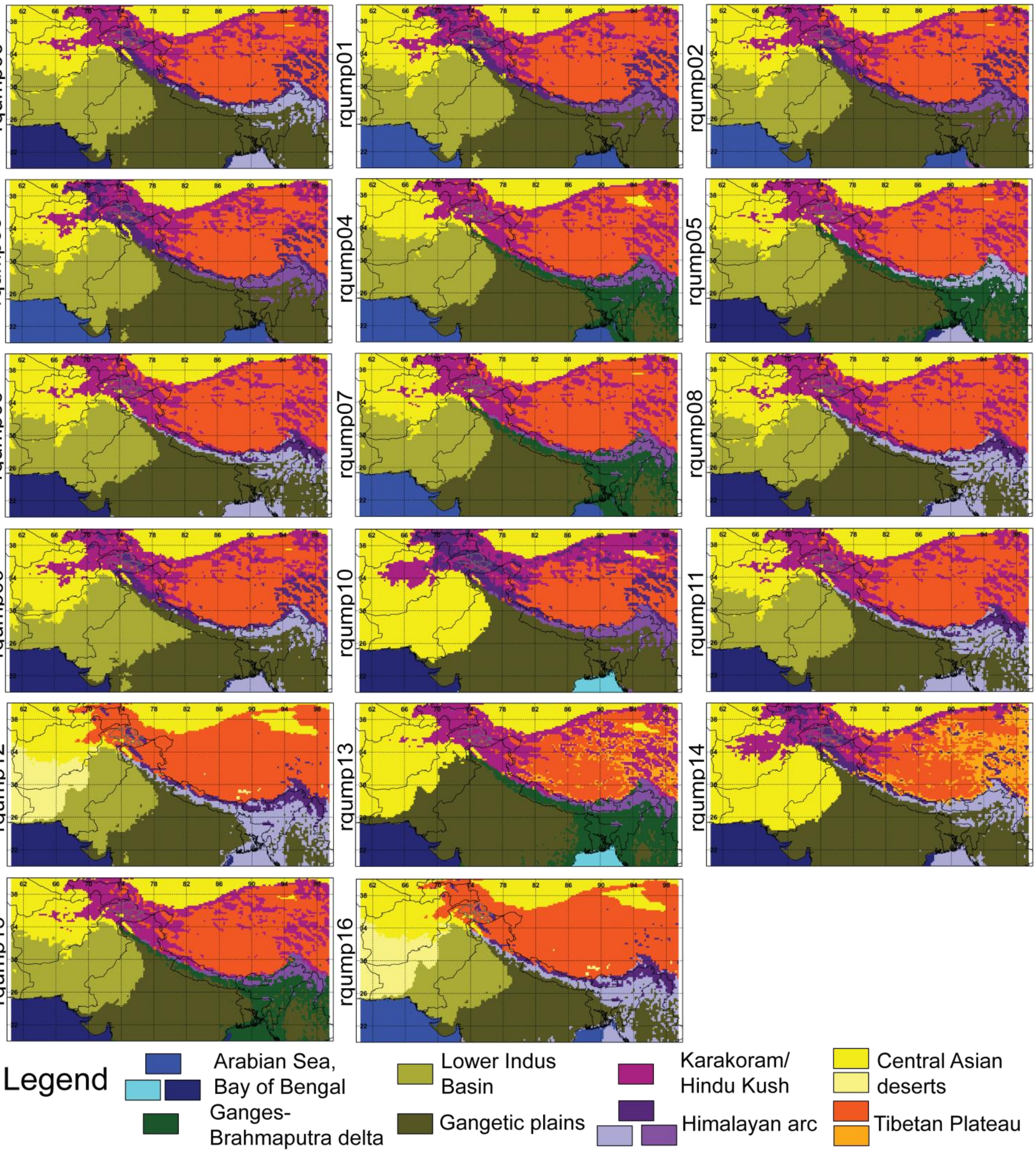

Figure 8. Comparison of climate classifications resulting from the use of 8 clusters on principal components of the control period (1970 to 1999) from the individual members of the Hadley Centre RQUMP perturbed physics ensemble downscaled over South Asia.