Kallio, Marko; Guillaume, Joseph; Kummu, Matti; Virrantaus, Kirsi-Kanerva

Spatial Variation in Seasonal Water Poverty Index for Laos: An Application of Geographically Weighted Principal Component Analysis

Published in:
Social Indicators Research

DOI:
10.1007/s11205-017-1819-6

Published: 01/12/2018

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Kallio, M., Guillaume, J., Kummu, M., & Virrantaus, K.-K. (2018). Spatial Variation in Seasonal Water Poverty Index for Laos: An Application of Geographically Weighted Principal Component Analysis. *Social Indicators Research*, *140*(3), 1131-1157. https://doi.org/10.1007/s11205-017-1819-6

Title: **Spatial Variation in Seasonal Water Poverty Index for Laos: An Application of Geographically Weighted Principal Component Analysis**

Abstract

Water poverty, defined as insufficient water of adequate quality to cover basic needs, is an issue that may manifest itself in multiple ways. Extreme seasonal variation in water availability, such as in Laos, located in Monsoon Asia, results in large differences in water poverty conditions between dry and wet seasons. In this study, seasonal Water Poverty Indices (WPI) are developed for 8215 villages in Laos. WPI is a multidimensional composite index integrating five dimensions of water: resource availability, access to safe water, capacity to manage the resource, its use and environmental requirements. Principal Component Analysis (PCA) and Geographically Weighted PCA (GWPCA) were used to examine drivers of water poverty and to derive different weighting schemes. Three major drivers were identified: poverty, commercial/subsistence agriculture and village location. The least water poor areas are located around the capital city and along the Mekong River Valley while the highest water poverty is found in sparsely populated mountainous areas. Wet season WPI is on average more than 12 index points higher than in the dry season, but in some villages monsoon rain does not improve the situation. The results indicate large spatial and temporal differences in WPI within Laos. In analysis of WPI components, a mean-variance scaled PCA is recommended due to its capacity for uncovering processes driving water poverty. Extending to GWPCA is recommended when information on local differences is of interest.

Keywords: Water Poverty Index; Geographically Weighted Principal Component Analysis; Monsoon; Water Poverty; spatio-temporal analysis; Laos

1 Introduction

Water poverty is a multifaceted problem which occurs when there is not enough water available to cover basic requirements. Although there are other factors relating to physical water scarcity – namely economic, institutional and political issues (Molle and Mollinga 2003) – physical water availability may be considered the major driver in water poverty especially in dry and arid regions with high population pressure. In Lao People's Democratic Republic (commonly known as Laos), located in Monsoon Asia, water availability difference between dry and wet seasons is extreme with in some cases more than 90% of the annual precipitation occurring during the wet season from May to November (Babel and Wahid 2009). The Monsoon region therefore faces very large variation in water availability throughout the year. With climate change projected to reduce precipitation in the dry season, to increase evaporation and and lengthen the dry spell (Beilfuss and Triet 2014), it is important to understand the dynamics of water poverty in the region.

The Water Poverty Index (WPI) is a holistic, multidimensional tool developed to address the issue of water poverty specifically. Its development arises from the need to incorporate physical water availability, environmental water needs and societal dimensions into an easy-to-use decision making tool (Sullivan 2002; Sullivan and Meigh 2007; Sullivan et al. 2003). It is a composite index which is unstandardized in the sense that the selection of indicators within the components are not predefined but are case specific. The five components included in WPI were developed through consultation with water managers, scientists and stakeholders (Sullivan and Meigh 2007). The components are a) Resources (RES), b) Access (ACC), c) Capacity (CAP), d) Use (USE) and e) Environment (ENV). RES typically consists of different indicators representing water availability (surface and groundwater), internal and external water flows and intra-annual variability. ACC measures the degree to which the local population has access to safe water for sanitation and irrigation and is commonly measured as the penetration of safe water in the population. CAP consists of indicators of water management capability, based on education, health and finance. USE describes the amount of water being utilized and its contribution to the economy. Finally, ENV attempts to capture the environmental quality and impact of water management. (Sullivan 2002; Sullivan and Meigh 2007; Sullivan et al. 2003)

The WPI has been used in many studies over several scales; at a national scale (e.g. Cho et al. 2010; Komnenic et al. 2009; Lawrence et al. 2002), regional scale (e.g. Heidecke 2006), and local scale (e.g. Garriga and Foguet 2010; Sullivan et al. 2003). In Monsoon Asia, WPI has been applied in only a few local studies: Ty et al (2010) assessed local water poverty in the dry season in Srepok River Basin in Cambodia-Vietnam, while Guppy (2014) looked at water poverty in five villages in Cambodia-Vietnam. Both studies conclude that the water poverty level in the areas investigated was high.

WPI has not previously been studied in Laos, the study area of this paper, apart from the international comparison by Lawrence et al (2002), which treated the country as a single unit. The study places Laos in the water poor half of the countries compared. The countries which score worse than Laos are mostly located in dry and arid regions or are also poor in other ways (including neighbouring Vietnam and Cambodia). The study, however, does not consider the seasonality of water resources nor the spatial variability within the country, which therefore remain knowledge gaps. To the authors' knowledge, seasonal differences in water poverty using WPI as a tool have been addressed only by Ty et al (2010). Tang and Feng (2016) and Zhang et al (2015) have shown that WPI components vary considerably between years. The extreme difference between monsoon dry and wet seasons is likely to cause large differences in some of the WPI components seasonally as well.

As WPI is a composite index consisting of five components, their weighting is an important issue, because the weights represent the *importance* of each component to the overall WPI. Many papers have discussed the selection of weights in a WPI context due to their arbitrariness and subjectiveness (e.g. Heidecke 2006; Molle and Mollinga 2003; Sullivan and Meigh 2007). There are many options for deciding how to weight the components: equal weights, expert opinion, stakeholder consultation or analytically derived weighting.

The computation of "objective", data-driven weights is of particular interest, because the computation process itself has the potential to provide new insights. Though these insights have typically not been deeply analysed, especially with regard to WPI, several authors have used principal component analysis (PCA) (e.g. Cho et al. 2010; Garriga and Foguet 2010; Jemmali and Matoussi 2013; Jemmali and Sullivan 2014; Pérez-Foguet and Garriga 2011). This method assigns weights based on the variances in the WPI components, which helps to highlight differences over time or space. Furthermore, PCA can be localized to a certain geographic extent using Geographically Weighted PCA (GWPCA). GWPCA allows studying of spatial heterogeneity of WPI components and their relationships. It was introduced in Fotheringham et al. (2002), and further developed by Charlton et al. (2010) and Harris et al. (2011, 2015). It has been used in a number of studies; Harris et al. (2014) used GWPCA to re-design a water quality sampling campaign, the same authors later applied GWPCA to inspect the spatial structure of a water quality dataset (Harris et al. 2015). Lloyd (2010) used it to inspect the socio-economic structure of the greater Dublin area while Wei et al. (2016) used GWPCA to assess multivariate spatial inequality. However, it has not been applied to multidimensional composite indices before.

The aim of this study is to investigate the spatial and seasonal heterogeneity of water poverty in Laos. Understanding is built up by using several data-driven weighting schemes of increasing sophistication, consisting of simple variance-based weights, PCA-derived weights, and local, GWPCA-derived weights. In doing so, this study demonstrates the usefulness of PCA and GWPCA in identifying processes governing water poverty, suggesting there is value in wider adoption in computation of social indicators.

Section 2 gives background information on PCA, and how it can be extended to account for spatial heterogeneity using GWPCA. Then, the study area is introduced and the development of seasonal WPI is shown in Section 3, including the different weighting schemes assessed. The fourth section provides the empirical results of the study, and finally, Sections 5 provides a discussion and conclusions.

2 Background

This section gives a description of PCA and GWPCA for interested readers. To understand how these methods can be used for weighting, the reader may jump directly to Section 3.

2.1 Principal Component Analysis

The PCA method is more than a hundred years old, and is one of the most widely used multivariate analysis methods. The idea of PCA is to reduce the number of variables in data while retaining as much information as possible. This is done by transforming interrelated variables into a new set of variables, principal components (PC), which are uncorrelated with each other (Jolliffe 2002).

PCs of a data matrix $\boldsymbol{X}$ with n observations and m variables can be found using the symmetrical variance-covariance matrix $\sum$ where the sum of the elements on the main diagonal is the total variance in $\boldsymbol{X}$. If the data matrix $\boldsymbol{X}$ is standardized to zero mean and unit variance, the values in the variance-covariance matrix $\sum$ are equal to the correlation matrix for $\boldsymbol{X}$ and the sum of the elements in the diagonal is equal to the number of variables m. Thus, from standard linear algebra it follows by eigendecomposition that

$$\boldsymbol{LVL^T} = \sum \qquad (1)$$

where $\boldsymbol{V}$ is a diagonal matrix of eigenvalues and $\boldsymbol{L}$ is a matrix of eigenvectors. $\boldsymbol{V}$ represents the variances and the column vectors of $\boldsymbol{L}$ represent the loadings of variables in the corresponding PC. PCs are commonly reported in decreasing order of eigenvalues and thus, the first PC accounts for most of the variance in the data, the second PC accounts for the most variance not accounted for by the first PC, and so on. Typically (but not always), the first few PCs explain the majority of the variance in the data while the last few describe a considerably lower amount (Harris et al. 2011; Jolliffe 2002). Joliffe (2002) provides a detailed explanation of the derivation of PCA and the numerous applications it is used in.

In order to achieve reduction in noise or to limit analysis to the most important drivers of change, only the first PCs are retained. It is customary to retain a number of PCs which satisfy a certain rule, for instance all components with eigenvalue higher than 1 (latent root criterion) or the number of components that explain more than a certain amount of variation. The number varies from study to study and from field to field; e.g. 60% in social sciences and up to 95% in earth sciences (Hair et al. 2006). This reduces the number of PCs, concentrating on major sources of variance, while retaining the majority of the information in the original data $\boldsymbol{X}$.

This article takes the perspective that the loadings of the derived PCs can be interpreted as processes that explain the variation in water poverty. Restricting the number of PCs focusses attention on specific processes.

The standard PCA above assumes that the variances and covariances of the data are constant. In a spatial context, this means that, despite it being possible to map the results of PCA (provided that the data has a locational reference), the variances and component loads remain constant over the entire area of analysis (Harris et al. 2011).

2.2 Geographically Weighted PCA

PCA can, however, be modified to account for first order spatial effects (spatial heterogeneity) by replacing it with a variant called Geographically Weighted PCA (GWPCA). This is a form of weighted PCA, which weights the sample points based on their geographical proximity to a point-of-interest (*u, v*) and performing standard PCA using the weighted sample (Demšar et al. 2013; Harris et al. 2011). The weighting in GWPCA effectively changes the scope of the PCA from global (the entire dataset) to the local geographic space. Geographically weighted (GW) eigenvalues and eigenvectors for location *(u,v)* may be obtained by computing the GW covariance matrix

$$\sum(u,v) = \boldsymbol{X}^T\boldsymbol{W}(u,v)\boldsymbol{X} = \boldsymbol{LVL}^T(u,v) \tag{2}$$

where $\boldsymbol{W}$ is a diagonal matrix of weights obtained by a kernel density function.

The kernel density function is used to address spatial autocorrelation; a measure of the relationship of some variable between nearby spatial units (Getis 2010). In other words, spatial autocorrelation describes whether values of nearby observations are more similar than distant observations. Statistical measures of spatial autocorrelation (e.g. Moran's Index) can be used, for instance, to identify spatial clusters, to test assumptions of spatial stationarity or heterogeneity or to test model mis-specification (Getis 2010). Geographical weighting of PCA addresses the spatial heterogeneity of variances and covariances in the dataset it is used to investigate by weighting observations using a kernel centered on the point of interest.

The kernel density function used can be for instance a "boxcar" (a moving window centered on the point of interest where all observations within a certain distance are given a weight 1, and all others excluded), a "Gaussian", "exponential", or "bi-square" (Brunsdon et al. 2002; Fotheringham et al. 2002). All of these weighting schemes are based on a bandwidth (BW), a geographical distance that defines the extent of "locality". Alternatively, it can be defined as adaptive, based on the number of nearest observation points to include in the analysis rather than a fixed distance.

The selection of BW may be done by an expert with sufficient domain knowledge, or it can be optimized by cross-validation. The cross-validation approach minimizes the sum of residual variance (the proportion of variance not explained by the retained PCs) for all locations in the data (Harris et al. 2011). The optimal BW ensures that the maximum amount of variance is explained by the pre-selected number of retained PCs. When PCs are interpreted as processes that explain variation, the BW describes the spatial extent of the combined processes, fit to the whole dataset rather than to each process individually.

The interpretation of GWPCA results is challenging due to the large amount of data produced by running the analysis for *n* observation locations: an *n x m* matrix of eigenvalues and an *n x m x m* matrix of eigenvectors (Fotheringham et al. 2002; Harris et al. 2011). Charlton et al (2010) suggest several visualization methods to aid the interpretation: a) to map the local eigenvalues (proportion of variance explained by the PCs), b) to map the loadings for each variable in each PC, c) to map the "winning variable" of each PC, and d) to use multivariate glyphs to represent the loadings.

3 Material and Methods

3.1 Case study area

Laos is a land locked country located between Cambodia, Thailand, Myanmar, China and Vietnam (see *Figure 1*). 80% of the total land area is classified as mountainous. Laotians are rural and poor with approximately 70% living in rural areas and approximately 30% living in poverty (The United Nations in Lao PDR 2015). The Southwest Monsoon dominates the climate and causes distinct dry and wet seasons with more than 85% of precipitation occurring between May and November (Babel and Wahid 2009; Mekong River Commission 2011). Several large rivers flow through Laos. In addition to the Mekong, major rivers include Nam Ou, Nam Ngum, Se Bang Fai, and Sekong. Human impacts on water resources are relatively low, albeit local hotspots of water pollution may be found (Babel and Wahid 2009; Mekong River Commission 2007). In addition, only 0.9% of the discharge is withdrawn for utilization of which nearly 99% is used in agriculture. Despite agriculture having a high share of water use, 90% of rice crops in Laos are rainfed and primarily practiced for sustenance (Babel and Wahid 2009). Based on the GWPCA analysis of this study (Section 4.3), Laos is divided into the four regions shown in *Figure 1*.

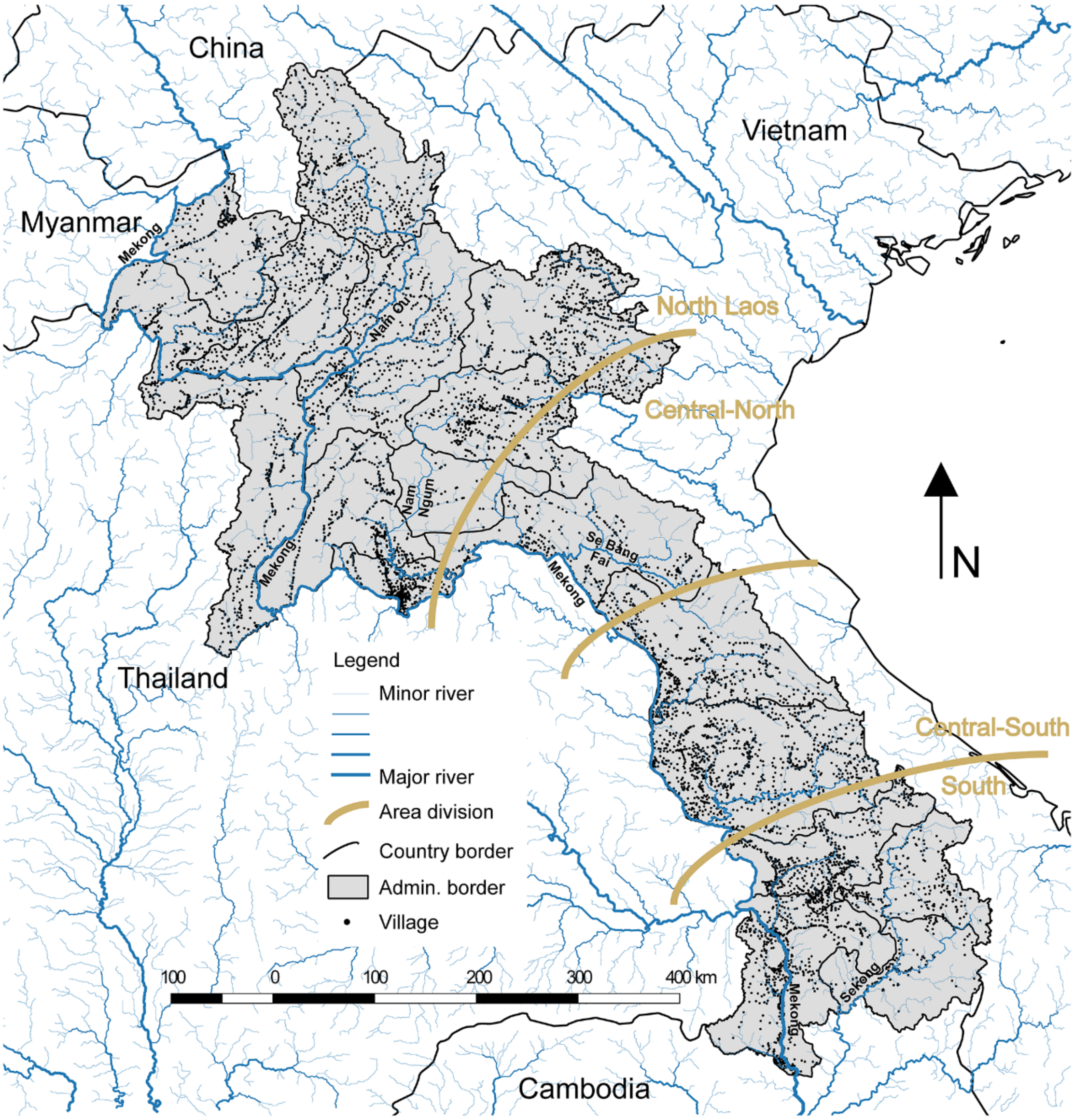

Figure 1. Map of Laos, showing the 8215 villages included in the analysis, overlaid with the approximate area division identified using the GWPCA analysis.

3.2 Developing the Water Poverty Index

WPI was calculated for 8215 villages in Laos, shown in *Figure 1*. These villages were selected because they are all represented in the two main data sources employed in this study: the Population Census of 2005 (Lao Statistics Bureau 2005) and Agricultural Census of 2010/2011 (Lao Statistics Bureau 2011). These data sources were complemented with hydrological modelling for water availability, observed precipitation from 183 stations across Mekong Region (data period 1981-2003), and SEDAC Last of the Wild v2 (Wildlife Conservation Society - WCS and Center for International Earth Science Information Network - CIESIN - Columbia University 2005) human footprint data, capturing human influence on terrestrial ecosystems. No new datasets were collected for the purpose of this study, following recommendations of Sullivan and Meigh (2007) to use pre-existing data as much as possible. The indicator selection for each WPI component is detailed below and summarized in Table 1. Statistical analysis was performed to ensure that PCA can meaningfully be applied to the components (according to the procedures in Cho et al. 2010; Hair et al. 2006; Hajkowicz 2006; Jemmali and Sullivan 2014) Details are given in Supplement A). Surface water availability and irrigation type were removed due to this analysis (Table 7 in Supplement A), and PCA was performed on the CAP indicators to create new uncorrelated variables. Low correlations were found between WPI components apart from ACC and CAP (Table 8 in Supplement A). WPI was calculated for dry season (dWPI, mid-November to mid-May) and wet season (wWPI, mid-May to mid-November) in order to capture the differences between them. Aggregation of indicators to a component score was done by averaging the indicators in their corresponding components.

Resources (RES)

Seasonal water availability was modelled using the VMod modelling software, which is a distributed physical 2-D hydrological model developed by Environmental Impact Assessment Finland Ltd (Koponen et al. 2010). The model was selected as it has proved to perform well in the Monsoon climate (e.g. Darby et al. 2016; Lauri et al. 2014; Räsänen et al. 2017). The model was applied using a grid with 5 km resolution. Input data about landuse, elevation, soil, temperature and precipitation was sourced from the Mekong River Commission and is the same as used in the baseline scenario in Lauri et al (2014). Detailed description of the model can be found in the model manual (Koponen et al. 2010).

For scoring, water availability per capita was used, following Ty et al (2010), rather than the total water availability. The threshold were taken from the Water Crowding Index (the Falkenmark Water Stress Indicator (Falkenmark et al. 1989)), giving a score of 100 to villages with water availability higher than 1700 m^3 $capita^{-1}$ $year^{-1}$ (no water shortage), and a score of 0 to villages below 500 m^3 $capita^{-1}$ $year^{-1}$ (severe water shortage).

Average daily precipitation was used to indicate the internal water resource and was scored as a ratio relative to the highest average precipitation in the wet season. The average longest seasonal period of consecutive days without precipitation (referred to as, consecutive dry days; CDD) was used to indicate how long, on average, the village is relying on external water flow. It was scored using the minimum length of CDD over both seasons (score 100) and the maximum length over both seasons (score 0).

Table 1. Indicators of the five WPI components, their scoring, data sources and whether the indicator value changes between dWPI and wWPI. RES is Resources, ACC is Access, CAP is Capacity, USE is Use and ENV is Environment components of WPI.

		Scoring		Does the variable change between seasons?	
Component	Variable	Minimum (0)	Maximum (100)		Data Source
RES	Surface water availability	<500 m^3/cap/year	>1700 m^3/cap/year	Yes	Simulated
	Average daily precipitation	0 mm	Max wet season precipitation	Yes	Lauri et al (2014)
	Annual longest consecutive drought days	Longest average annual dry period in days	Shortest average annual dry period in days	Yes	Lauri et al (2014)
ACC	Irrigation type	No irrigation facilities	Maximum number of different irrigation methods	No	Agricultural Census 2010/2011
	Drinking water source(s)	No specified source	Piped water	No	Agricultural Census 2010/2011
	Toilet type	No toilet	Modern	No	Population Census 2005
CAP	Travel time to province and district capitals	>600 min	Travel time 0 min	No	Population Census 2005
	Village road access	No	Yes	Yes	Agricultural Census 2010/2011
	Literacy rate	0%	100%	No	Population Census 2005
	Incidence of poverty	100%	0%	No	Population Census 2005
USE	Share of irrigated crops from total crop area	0%	100%	Yes	Agricultural Census 2010/2011
	Share of population depending on aqua- or agriculture for their income	100%	0%	No	Agricultural Census 2010/2011
ENV	Disaster occurrence	No disasters	All disaster types occurring every 1-2 years	Yes	Agricultural Census 2010/2011
	Human Footprint	100	0	No	SEDAC Last of the Wild v2
	Soil degradation	Severe degradation	No degradation	No	Agricultural Census 2010/2011

Access (ACC)

This study uses the presence of water infrastructure as a proxy for the penetration of safe water, which is missing from the census data used. The selected indicators are the presence of infrastructure for irrigation (permanent weir, reservoir, pump scheme, gates/dykes, temporary weir, gabions or other irrigation scheme), main drinking water source (piped water, protected well, unprotected well, surface water; mountain water source, rain water or other) and main toilet type (modern, normal, other or no toilet). Irrigation facilities are given the highest score for villages with all different irrigation methods, and zero score for those with no irrigation facilities present. Scoring for main water source and toilet type are given in Table 2.

Table 2. Scoring of Access indicators

Indicator	Type	Score
Main water source	Piped water	100
	Protected well	83
	Unprotected well	67
	Surface water	50
	Mountain water source	33
	Rain water	17
	Other	0
Toilet type	Modern	100
	Normal	66
	Other	33
	No toilet	0

Capacity (CAP)

The Capacity component measures the capability to manage available water resources. In this study, four indicators are used as proxies of this; a) combined travel time to district and provincial capital (services available in each differ), b) road access, c) literacy rate, and d) incidence of poverty. Travel time to the capital(s) is scored 100 if the village is the provincial and district capital (travel time 0 minutes), and 0 if the combined travel time is above 10 hours (600 minutes). The ten-hour threshold is an arbitrary choice, but is selected assuming that it gives a meaningful indication that either capital may be visited within two days. In addition, it should be noted that the travel time likely changes throughout the year, but data on the effect of season was not available in the dataset used. Village road access is given a score of 0 for villages without, and score of 100 for villages with road access. The census data shows that the road access varies according to season; in some cases, roads are blocked in the wet season due to flooding. Literacy rate is provided in the census by percentage of total village population that can read, and it is taken directly as the score. Incidence of poverty is likewise provided as percentage of total village population, and the score is computed by subtracting it from 100, which gives higher score to villages with low poverty.

Use (USE)

Use typically includes information about the quantity of water used by different water use sectors. This type of data, however, was not available from the data sources and therefore two indicators were used as proxies: a) the percentage of crops being irrigated from the total crop area, and b) the percentage of population (from the total village population) depending on agri- or aquaculture for their income. The share of crops irrigated is taken as is. For the dependency on water, the percentage is subtracted from 100 in order to give higher score to those villages with lower income dependency on water. Of these two, share of irrigated crops varies according to season. It is noteworthy that due to the selection of these two indicators, in this study USE is interpreted as a measure of vulnerability.

Environment (ENV)

Several studies use different indices to represent Environment, such as water quality index. However, no such data exists in high enough resolution from Laos, and thus, proxies were used. The indicators include a) soil degradation (from no degradation to light, moderate or severe degradation), b) disaster occurrence (drought, flood, landslide, pests, other, non-specified) (Lao Statistics Bureau 2011), and c) human footprint from SEDAC Last of the Wild v2 dataset (Wildlife Conservation Society - WCS and Center for International Earth Science Information Network - CIESIN - Columbia University 2005). It is assumed that droughts occur only during dry season, and that floods and landslides occur solely during wet season making disaster occurrence vary between seasons. Scoring for disaster occurrence is counted so that if there are no disasters occurring in a village, the score is 100, and having all disaster types occurring every 1-2 years, the score is 0. Human footprint in the SEDAC dataset is a number between 0 (natural environment) and 100 (entirely human modified environment). It is assumed here that the lower the human footprint, the higher is the integrity of the ecosystem. Thus, the human footprint value is subtracted from 100 to give higher score to villages with lower footprint.

3.3 Case study WPI component weighting schemes

Five weighting schemes used to aggregate the components to a WPI were derived and compared. This section describes reasoning behind the progression of weights from variance-based weights to local, GWPCA derived weights. As a base case, the first weighting scheme assigns equal weights to each component.

The remaining weighting schemes are based in some way on PCA. The approach used is based on existing literature (e.g. Cho et al. 2010; Jemmali and Matoussi 2013; Jemmali and Sullivan 2014). In PCA-based weighting, the weights are computed from eigenvalues ***V*** (the variances) and eigenvectors ***L*** (the component loads) of a number of retained PCs. Since PCA is based on the variance of the variables in data (in this case, the variance of WPI components), this approach gives higher weight to the WPI components which are highly loaded and lower weight to those which are weakly loaded in the first few PCs, emphasising the variables that contribute most to the variance of the WPI components. The PCA derived weights can be computed by multiplying the squared component loads and the proportion of variance explained by the corresponding PC, and summing across PCs. Weights are therefore derived using Eq. 3

$$\beta_i = \sum_{k=1\dots n} PC_{k,i}^2 \times \frac{\sqrt{\lambda_k}}{\sum_{j=1\dots n}\sqrt{\lambda_j}} \tag{3}$$

where β_i is the weight given to WPI component *i* (either RES, ACC, CAP, USE or ENV), PC_{ki} is the component load in *k*th PC (column of ***L***), λ_k is the eigenvalue of the *k*th PC (in ***V***) and *j* is the number of PCs retained. This method of weighting is a modified version of that used by Jemmali and Sullivan (2014) with a difference that here the component load is squared. This has the advantage that the squared sum of an eigenvector is 1, such that the sum of squared principal component loads for each of the WPI components corresponds to the proportion of its variance relative to the full dataset (Jolliffe 2002). Compared to Jemmali and Sullivan's (2014) method, this avoids obtaining negative weights β_i, cancelling out of PCs due to opposite signs, and requiring further scaling for the components to add to one.

The method of aggregation of components is also based on existing literature. Gariga and Foguet (2010) find that the classic style of using additive function to compute the final WPI suffers from full compensability, meaning that a low score in one component may be completely offset by a high score in another component. They instead recommend a weighted geometric mean:

$$WPI = \prod_i x_i^{\beta_i} \tag{4}$$

where x_i is the score of WPI component *i*. This prevents compensability of components in the WPI and increases the sensitivity to variation in each component. In this study, all component scores <0.01 were set at 0.01, in order to avoid the geometric mean setting the total WPI to zero because of a single component with value zero.

Since weights are based on eigenvectors and eigenvalues from PCA, GWPCA can be used to derive weights for each individual location based on geographical proximity. This can be done by replacing the terms β_i, PC_{ki} and λ_k in Eq. 3 with their local versions $\beta_i(u,v)$, $PC_{ki}\,(u,v)$ and $\lambda_k(u,v)$, as derived in Section 2.2.

The second weighting scheme is selected to provide a foundation for understanding the behavior of PCA-based weighting – guiding the interpretation of the retained PCs as potential "processes", as proposed in this study. It provides an intermediate step between the first, equal weight, scheme, and true PCA-based weighting. Given that the method of calculating weights based on PCA gives higher importance to the components with higher variance in their values, it follows then that, if there is a substantial difference in variances, the components with higher variance will dominate the retained higher variance principal components. To help understand this effect, weights are derived from variances of the WPI components, i.e. variance of the component/sum of the variances.

Because there are large differences in variances of WPI components (see *Table 3*), the third and fourth weighting schemes were respectively computed using PCA first without, and then with mean-variance standardized components. Mean-variance standardization scales the components to mean of zero and variance of one. According to Chatfield and Collins (1980, cited in Harris et al (2011)), this can be thought as giving each component equal importance in the analysis. The effect on weights is that none of the components dominates through having a higher variance. Note that the first two weighting schemes, using equal weighting and variance-based weighting, respectively correspond to the special cases of PCA-based weighting retaining all PCs with mean-variance scaled, and unscaled indicators.

GWPCA is used for the fifth weighting scheme. Its use is motivated here by the observation that the indicators are strongly spatially autocorrelated according to the commonly used Moran's I values, which are all >0.30 (with the exception of soil degradation, with 0.26). Visual inspection of the mapped variances also reveal large differences in their spatial distribution. It may therefore be expected that weights computed from the local results would vary spatially. Bandwidth selection for GWPCA was optimized using the cross-validation approach comparing boxcar, exponential, gaussian and bi-square kernel functions. An adaptive bandwidth was chosen because it a) provides the same number of sample points for each location and b) it addresses the problem of edge effects (Brunsdon et al. 2002).

4 Results

This section first gives an overview of the WPI component values and WPI computed using non-PCA methods (equal and variance based weights). Interpretation of results focusses on spatial variation of WPI components Then, PCA is applied to the components and its results discussed along with PCA weights, highlighting socio-economic processes driving water poverty. The weights used to aggregate the components into WPI are shown in *Table 3*. Finally, GWPCA is performed and the spatial heterogeneity of components and weights is investigated, yielding a local analysis that reveals hidden processes.

4.1 Spatial variation of WPI components

The WPI components vary greatly in their scores. To distinguish between dry and wet season scores, WPI and its components are prefixed with "d*" or "w*" for dry and wet season respectively when a season specific WPI or its component is referred to. In the dry season, lowest mean scores are for dRES, 21.6, and for dUSE, 26.5 while the highest mean scores are found for dCAP, 76.5 and for dENV, 78.5. dACC falls in between with a mean score of 43.3 (the higher the score, the lower the water poverty). The wet season changes the scores in all but ACC, since it does not have variables that change according to season. wRES increases, on average, 50 points from dRES, rising to 72.3. It is noteworthy that all of the villages have wRES at a minimum 10 index points higher than their corresponding dRES. The other components increase a considerably smaller amount. wUSE (30.6) is on average 5.1 points higher than dUSE. Increase in ENV is negligible, being on average 0.3 index points. CAP is the only component which, under these indicators, scores lower in wet than in dry season; on average the decrease is 11.5 index points (dCAP 76.5 vs. wCAP 65.0). The component values suggest that issues in village-level water poverty are mostly in access to safe water (due to missing infrastructure) and in USE component (in this case study, income vulnerability). In addition, RES scores suggest that in the dry season villages are reliant on external water flows.

The spatial and temporal patterns of the WPI components are mapped in Figure 2 using a geographically weighted mean (see Brunsdon et al., 2002) of 50 nearest villages. Clear patterns are visible in the components. dRES scores highest in the northwestern part of the country and lowest in the central-south. This is because the difference between dry and wet season precipitation is the lowest in the northwest. For ACC, the area around the capital city scores high, while the rest of the country scores significantly lower. The pattern in CAP shows that the mountainous areas bordering Vietnam and China are of considerably lower capacity than the Mekong River Valley bordering Thailand. In USE, clear hotspots of very low scores (high vulnerability) can be found in several areas in the north as well as in the Bolaven Plateau in the south. USE scores in the wet season increase mostly in the North in a few areas, while there is no or very little increase in the capital area and in the south. Finally, ENV scores high all throughout the country and shows only very small changes both spatially and temporally.

The spatial patterns in individual components are brought to various extents in the aggregated result depending on the weights used. Variances in *Table 3* (column 2) show that components which describe the environment (RES, ENV) vary considerably less than the "human" components. The interpretation here is that socio-economic factors cause more variation in intra-seasonal WPI in Laos than the environmental ones do and hence, are weighted accordingly. However, when the temporal scale is shifted to the annual level, combining data from both seasons, the variances change. Resource availability becomes as important as CAP in terms of variance and variance-based weights. This then means that overall WPI computed for seasons separately or combined respectively have similar and different distributions (*Table 4*). RES is the component that changes most between dry and wet seasons, but receives a low weight. These results correspond to expectation regarding the important role of water availability variation temporally, between seasons, but less important role spatially, given the relative spatial homogeneity of the monsoon. It should also be noted, that the monsoon rains are, in addition to measures of water availability, the ultimate reason in variation of CAP as road access is hindered due to excess water.

Table 3. Unscaled and scaled variances and the global weighting schemes used in the study. Equal, variance based and PCA based weighting schemes are given for dry season, wet season and both season data. Equal weighting scheme is equivalent to PCA derived weights when scaled data is used and all principal components are retained. Variance weighting scheme is equivalent to PCA derived weights when unscaled data is used and all principal components (PC) are retained.

	Component	Unscaled Variance	Scaled Variance	Equal weights / Scaled PCA weights using all PCs	Variance weights / Unscaled PCA weights using all PCs	Unscaled PCA weights	Scaled PCA weights
DRY SEASON	RES	96.3	1	0.2	0.064	0.005	0.127
	ACC	499.3	1	0.2	0.334	0.365	0.198
	CAP	423.9	1	0.2	0.284	0.284	0.212
	USE	389.4	1	0.2	0.260	0.346	0.220
	ENV	86.2	1	0.2	0.058	0.000	0.243
WET SEASON	RES	49.6	1	0.2	0.024	0.000	0.260
	ACC	499.3	1	0.2	0.243	0.187	0.206
	CAP	992.8	1	0.2	0.483	0.579	0.201
	USE	420.4	1	0.2	0.205	0.234	0.110
	ENV	91.3	1	0.2	0.044	0.000	0.224
BOTH SEASONS	RES	714.8	1	0.2	0.291	0.338	0.184
	ACC	499.3	1	0.2	0.204	0.161	0.195
	CAP	741.3	1	0.2	0.302	0.320	0.201
	USE	409.2	1	0.2	0.167	0.181	0.160
	ENV	88.8	1	0.2	0.036	0.000	0.259

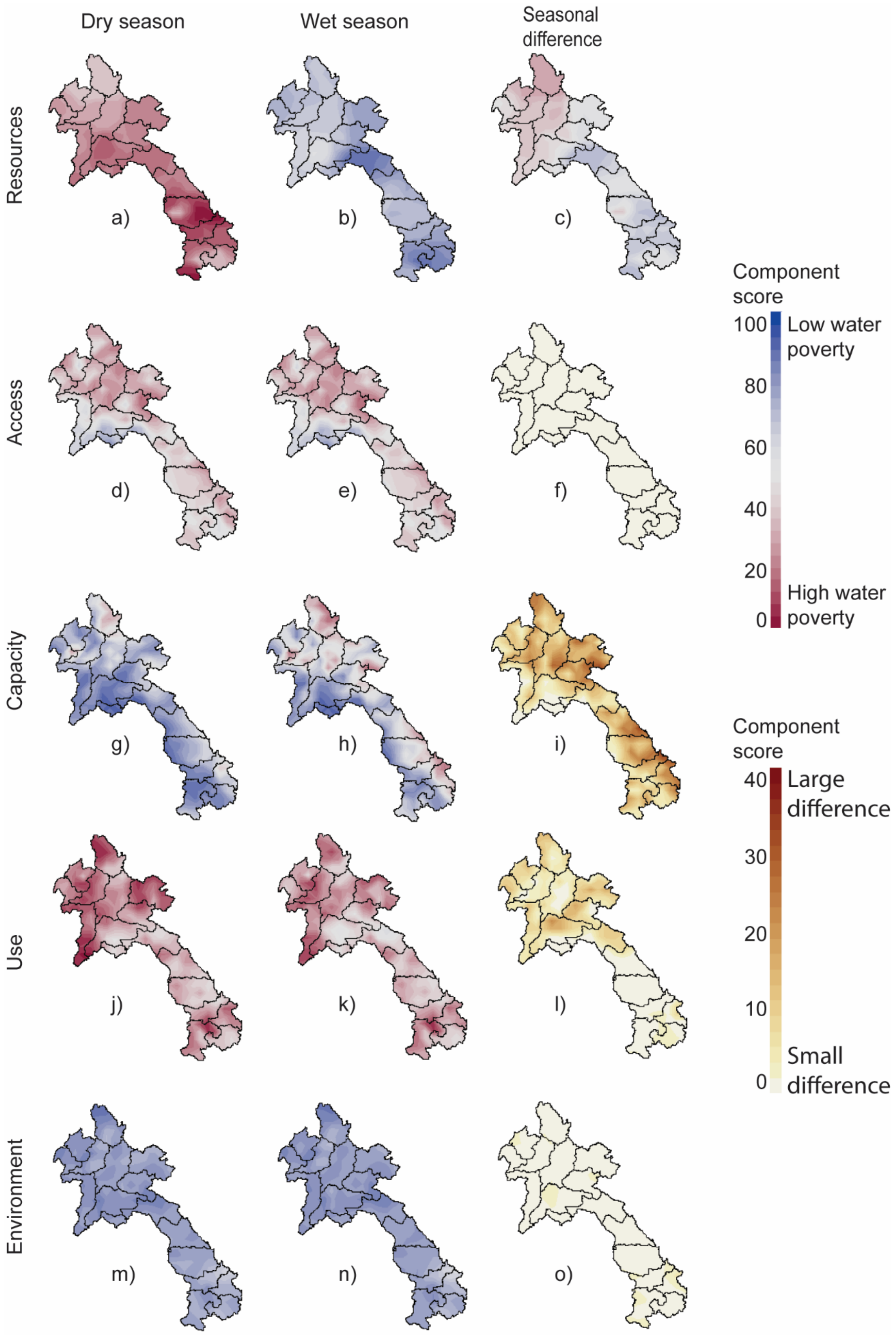

Figure 2. Seasonal comparison of WPI components shown as local geographical weighted mean of 50 nearest villages using boxcar weighting scheme, showing dry season in the first and wet season in the second column. The (absolute) difference between dry and wet seasons are given in the third column. High score represents low water poverty. Note: the tiles within the box (i, l and o) have different scale to other tiles. For the colour reference, see the online version.

While the ranges and mean values of WPI differ from one weighting scheme to another, the spatial distribution of higher and lower WPI areas is similar, as can be seen by comparing the maps in Figure 3. The highest scoring areas are found around the capital city Vientiane and in a small area between Bolaven plateau (which is visible as the round area of low scores in the very south) and Cambodian border. The Mekong River valley likewise score high near the Thai border. Lower scores are found in the mountainous and remote areas near the Vietnamese border as well as in large parts of northern Laos. In some cases, provincial borders can be identified from the WPI values, especially in the South around Bolaven Plateau and in central Laos between Bolikhamxai and Khammouane.

Table 4. Summary statistics of WPI computed from equal weights, variance weights based on data from individual season, and variance weights based on both season data.

	Equal weights		Variance weights (single season)		Variance weights (both seasons)	
	dWPI	wWPI	dWPI	wWPI	dWPI	wWPI
Min.	1.57	1.34	0.31	0.44	1.74	1.36
1st Quartile	27.99	36.71	24.82	27.64	24.78	35.34
Median	37.11	49.58	39.02	46.54	33.75	50.26
Mean	36.64	48.70	37.97	45.66	33.65	49.00
3rd Quartile	46.76	61.87	50.23	62.88	43.08	63.18
Max.	70.11	86.99	86.64	94.75	65.05	88.71

Visual inspection of the maps in Figure 3 reveals that the spatial dynamics are comparable – similar kind of visual interpretations can be made. In addition to the spatial distribution, the difference between seasons is clearly seen. However, the extent to which water poverty decreases in the wet season varies by location. Generally, the situation improves the most in the areas which are least water poor in the dry season. The areas which are in the most difficult situation receive the least relief as the wet season sets in. The effect seems to be driven by the combined changes in CAP and RES: the increased water availability in villages in the poorest situation is offset by cut off road access in the wet season. This is attributed to general poverty levels and infrastructure.

Overall, the variances and weights based on them seem to suggest (bearing in mind that water availability was not an issue in majority of the villages) that in Laos water poverty is mostly driven by the “human components” and thus is more a management problem.

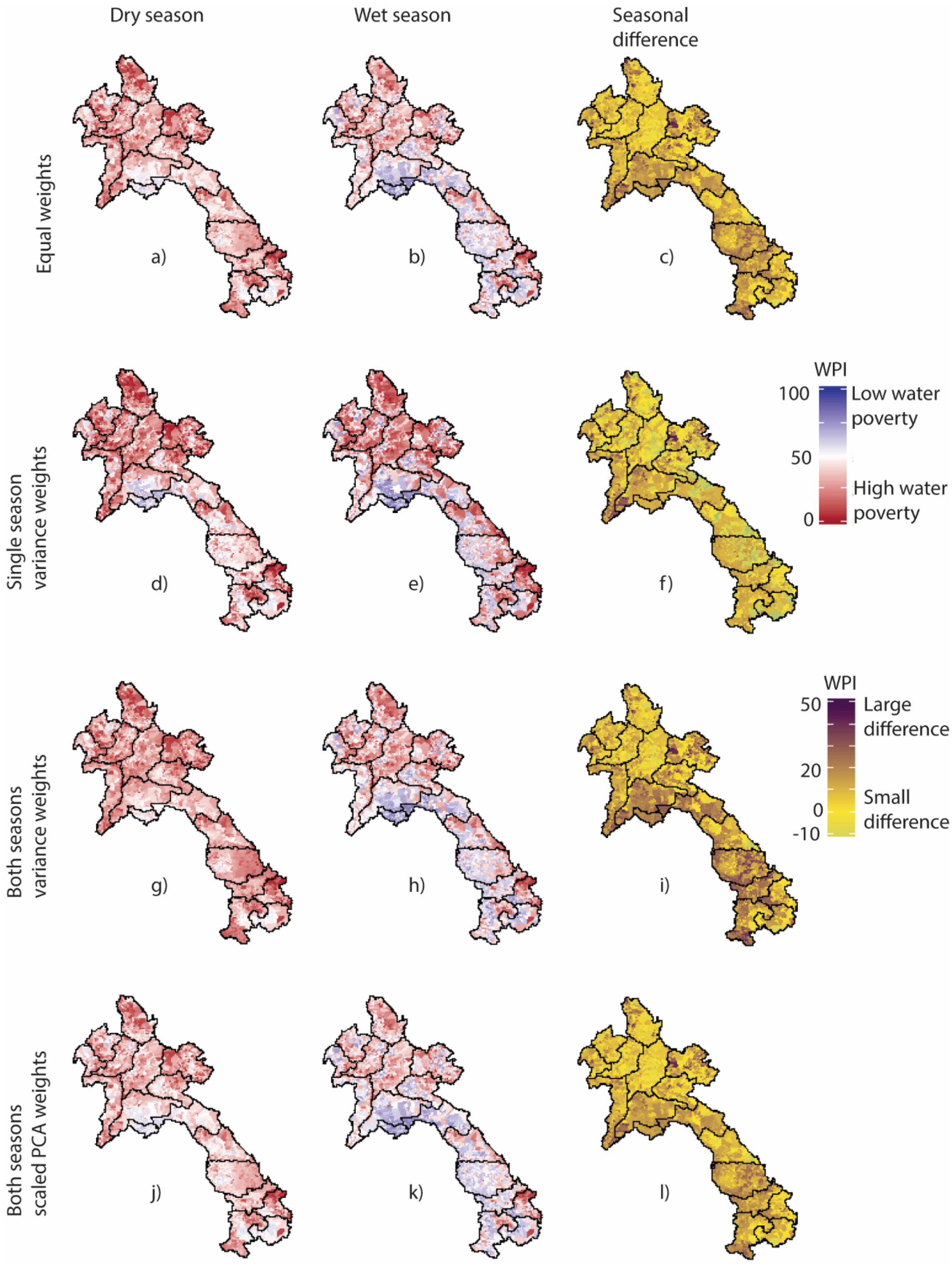

Figure 3. WPI outcomes from equal, single season variance, both season variance and both season PCA derived weighting schemes and the seasonal difference between them. A high WPI score represents low water poverty where as low score represents high water poverty. Note: Seasonal difference has different colour scale from the dry and wet season columns. For the colour reference, see the online version.

4.2 *Socio-economic processes drive water poverty*

Moving from a variance-based weighting into analytical method using PCA provides new interpretations regarding the causes of variation in WPI in Laos. Running a PCA analysis on the unscaled, raw WPI components yields two retained components for dry and wet seasons, explaining

73.3% and 79.8% of total variation respectively. The retained and excluded PCs are shown in *Figure 4*. As mentioned previously, the PCs are interpreted as processes that explain the most variance in the data. The unscaled, single season PCA return identical retained processes for both seasons: approximately half of all variation is explained by the “human” process with ACC, CAP and USE highly loaded, and where RES or ENV are not involved. The second PC, or process, is one where vulnerability (USE) decreases as CAP increases. The interpretation is that the most important driver of water poverty in Laos is poverty (state of development) in general, and the dependence on water (agriculture) of the rural population.

PCA results from both seasons combined, however, result in three retained PCs, the first of which explains 41.7% of variation, and contrasts abundant resources with low ACC and CAP. This could be interpreted as rural villages being located in locations with reliable water resources (higher precipitation, shorter dry periods). The second, PC2, describes a situation where increased water availability is coupled with increased infrastructure and increased income vulnerability. The third retained process, PC3, involves strongly increased dependency on agricultural products with increased capacity and water availability. This is assumed as distinguishing the extent to which villages engage in commercial rather than subsistence farming.

All of the above PCs make sense in light of the variances of components. All single season PCs are strongly loaded with the human components that show large variance. In the case combining seasons, RES is included, due to its large variance. However, scaling the raw component scores prior to PCA reveals richer, more detailed interactions between WPI components. The reason is that with mean-variance scaled variables, PCA is based on *interactions* (correlations) between the WPI components rather than the magnitude of variance.

Three PCs are retained in each of the scaled PCA analyses, explaining from 72.9% to 74.5% of total variance. The first PC in each (scaled dry, wet and both seasons) is a similar process to the most important process in the unscaled analyses: the human components contrasted to the environmental components with the only difference being that in dry season, RES is a significant factor while it is not with the others. This process accounts for a third of the overall variance. The second and third PC, however, differ between seasonal weighting schemes.

In the dry season, dPC2 is a process where income vulnerability increases as the score in all the other components increase. The same process is present in the wet season as the third most important process (wPC3). The interpretation here is that as the villages have better access to infrastructure, higher capacity and more water, they move from subsistence farming towards commercial agriculture. The third process in the dry season (dPC3) is one where USE and ENV increase with a modest decrease in CAP, ACC or RES, which suggests that rural villages (low human impact, low soil degradation) engage in subsistence rather than commercial agriculture. In the wet season, wPC2 describes the extent to which village locations having high internal water resources is accompanied by a decreased ENV score (higher disaster occurrence, soil degradation).

Annual scale changes the identified processes. The second most important process is represented by increasing USE wth increasing RES, suggesting that higher internal water resources lead to higher share of irrigated farmland. Finally, the third retained PC is a process where ENV is represented alone, which is a function mainly of village location: whether rural or urban, or higher or lower disaster occurrence.

Overall, the scaled PCA analysis reveals processes that could not be found with unscaled PCA by the inclusion of environmental components in addition to the human ones. Both analyses, however, agree that ACC, CAP and USE are the main drivers of water poverty, and is here interpreted as general state of development of the villages. The processes also put an emphasis on village, which is linked to the urban/rural dimension (i.e. access to markets, access to administration).

The retained PCs described above translated to weights are shown in *Table 3*. PCA-derived weights for the unscaled WPI components are very close to the variance based weights presented in the previous section with the stark difference that RES and ENV are given effectively zero weight for dry and wet seasons. RES is shown as the most important component when both seasons are used, while ENV does not receive any weight. This reflects the retained PCs. PCA derived weights based on scaled components, reflected by the difference in PCs in Figure 4, are entirely different; ENV becomes the most important component. This highlights the effect of scaling of components; unscaled components miss the important spatial aspect by neglecting the location of the village. The weighting schemes overall are more balanced and near the equal weighting scheme where each component receives weight 0.2. This is due to low inter-component correlations in the case study components with only CAP and ACC being strongly correlated.

WPI calculated from scaled both-season weights results in similar visual interpretation to variance based weights; the same areas are highlighted as more and less water poor. The mean difference between dWPI and wWPI is 12.3 index points. Interestingly, in a large number of villages dWPI and wWPI are within two index points, indicating that there are areas, mostly in Northern Laos, where increasing water availability in the wet season does not improve the water poverty situation in the villages. This further highlights the fact that water poverty is a socio-economic problem.

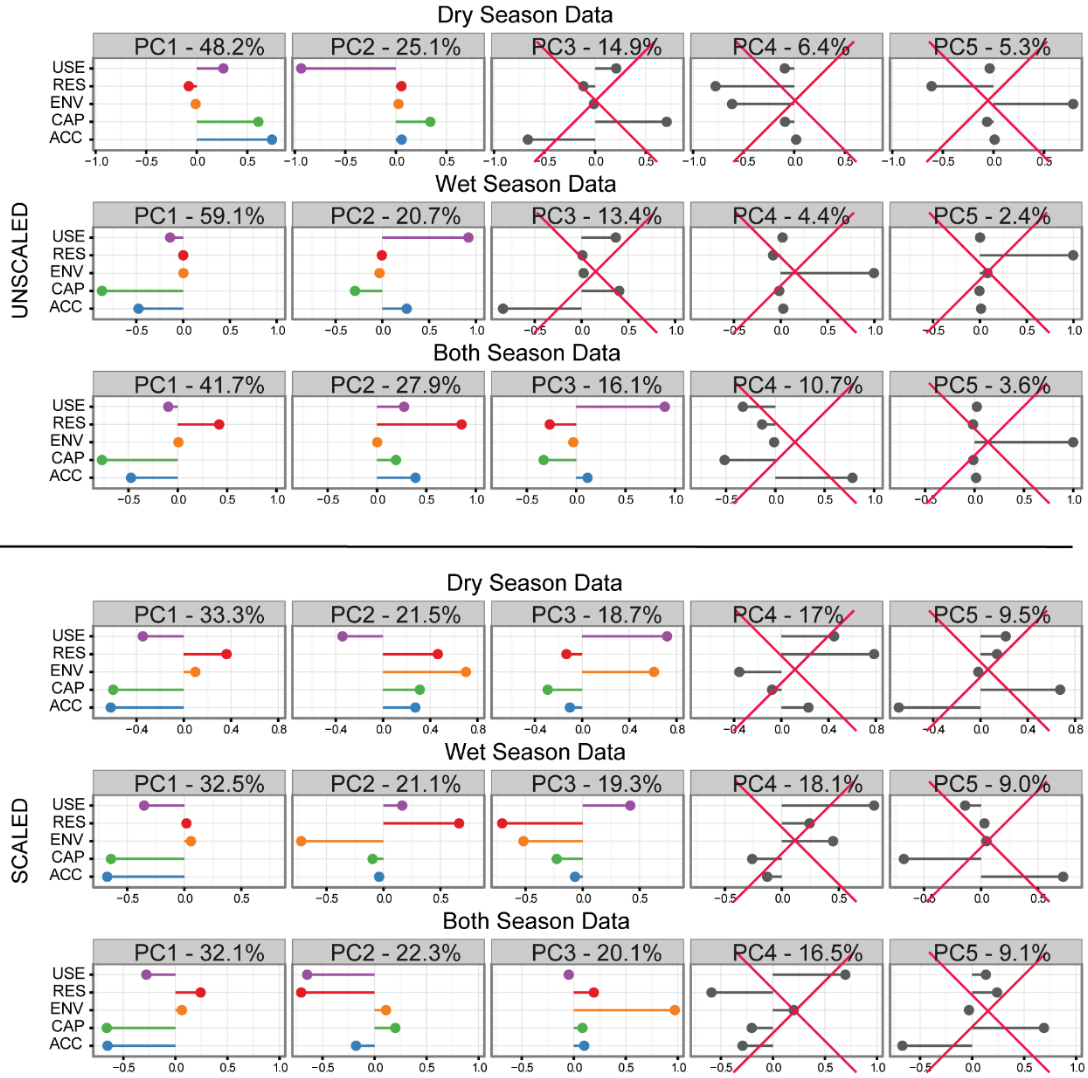

Figure 4. Unscaled and scaled PCA component loadings for dry, wet, and both season data. The crossed out Principal Components were discarded according to the used rule of strictly retaining the number of PCs that explain more than 70% of total variance.

4.3 Local analysis reveals hidden processes

The PCA results above describe global (country-wide) processes that explain the general trends within Laos. What about spatial differences in the processes? This section provides results on the spatial extent, the bandwidth, of the *local* processes, what they are, and how local weights can contribute to interpretation of the WPI components.

The bandwidth (BW) used for weighting of villages is an indicator of the strength of the extent of spatial vs. global process(es). *Table 5* provides the cross-validation optimized BWs using un-scaled and scaled WPI component scores. There is little difference between dry and wet season adaptive BWs at approximately one third of all villages. This is a large BW (several hundred kilometres, depending on the location of the village), but it is statistically significant – there is local spatial variation. Dry season BW using scaled data is similar to the unscaled. However, in the wet season BW drops to 680 villages, and the optimal function changes from boxcar to exponential. Thus, the spatial dimension of the local processes is larger in the dry season than in the more varying environmental conditions of the wet season.

Table 5. Optimal cross-validated adaptive bandwidths for scaled and unscaled component data. All of the bandwidths are significant with 0.01 confidence level under Monte-Carlo eigenvalue randomization test.

	Bandwidth	Function
Unscaled		
Dry season	2810	boxcar
Wet season	2500	boxcar
Scaled		
Dry season	2710	boxcar
Wet season	680	exponential

The PCs were interpreted as processes as in the global analysis, and results of scaled GWPCA analysis are shown in *Table 6*. Maps of the components and their associated loads are given in *Figure 7* and *Figure 8* in Supplement B to this article. The three retained local PCs explain between 73.6% and 88.4% of the local variance in the dry season and between 70.9% and 83.3% in the wet season, depending on location. In interpreting the GWPCA, only the components with higher than 0.3 load are included.

Similar and different processes can be found in the local analysis as was found in the global one in Section 4.2. The most important processes (PC1) in north and central-north Laos, on both dry and wet seasons, is the same poverty related process involving the human components ACC, CAP and USE. In the same area, the second most important process (PC2) is related to either ENV alone (dry and wet season, the location based process identified in global analysis as well), RES alone (wet season, this was not seen in the global analysis) or ENV and USE together (dry season) with a positive relationship. The third identified process in the north and central-north is related to income vulnerability (USE). However, there are important differences: in the north, dUSE and dACC have a positive relationship (vulnerability decreases as safe water infrastructure is present), whereas in in central-north Laos, the relationship is opposite. These opposite relationships are interpreted as the northern villages being dominated by subsistence farming (less dependent on water for their income), while in central-north along the fertile river valley, villages engage more in commercial farming. In central-north Laos, a negative relationship exists also between dUSE and dENV, which suggests that commercial farming is practiced in more developed areas rather than in the rural villages. In addition, both north and central-north areas exhibit also a negative relationship between dCAP and dUSE. In the wet season, the last identified process (wPC3) is characterized by increasing wUSE and wENV. The process is interpreted as opposite from the dry season: decreasing environmental pressure and decreasing income vulnerability, which may be due to rural villages engaging in subsistence farming.

Interpretation of the processes in central-south and south is more challenging due to many subareas rather than general trend in the components. The central-south may be divided into two parts; riverbanks of Mekong, and the “inland” of the country, to the Vietnamese border. In the most important process of the “Mekong area”, dUSE and dRES contrast dENV while in the “inland” area dUSE and dCAP contrast dENV. These are interpreted as describing agricultural processes; the more developed (lower environmental integrity) the village is, the higher proportion of population engage in commercial farming. The second and third process in central-south is a mix of different components contrasting and supporting one another, and where generalizations cannot be made. However, it does seem that there are processes that may counterbalance each other in the “Mekong” area, since dUSE and dENV show both negative and positive relationships in different PCs.

Table 6. Results of GWPCA for the three retained PCs and the identified processes. The processes are broken down to four regions with each process reported under that region. Red color signify positive relationship with the lead item (the highest loaded component, always positive), while blue colour signifies negative relationship. The components are marked mixed if in the region and in the process there are too many involved to identify a representative component. See division of regions in Figure 1.

Season	Component	Area Process	North			Central-North			Central-South			South		
			Lead	Second	Third	Lead	Second	Third	Lead	Second	Third	Lead	Second	Third
DRY SEASON	PC1	1	dACC	dCAP	-	dACC	dUSE	dCAP	dENV	dUSE	dCAP	dRES	dCAP	-
		2	dCAP	dACC	dUSE				dUSE	dENV	dRES			
	PC2	1	dENV	-	-	dENV	dUSE	-	mixed	mixed	mixed	dENV	dUSE	-
	PC3	1	dUSE	dACC	-	dUSE	dENV	dACC	dUSE	dENV	mixed	dUSE	dENV	-
		2	dUSE	dCAP	-	dUSE	dENV	dCAP	mixed	mixed	mixed			
WET SEASON	PC1	1	wCAP	wACC	-	wACC	wCAP	wUSE	wCAP	wACC	-	wENV	wCAP	mixed
		2	wACC	wCAP	wUSE				wCAP	wENV	wUSE	mixed	mixed	mixed
	PC2	1	wENV	-	-	wRES	-	-	wENV	wCAP	wACC	mixed	mixed	mixed
		2	wRES	-					wUSE	mixed	mixed			
	PC3	1	wUSE	wCAP	-	wENV	wUSE	-	wUSE	wENV	-	wUSE	wENV	wRES
		2	wUSE	wACC	wENV				wENV	mixed	mixed			
		3	wUSE	wENV	-									

Similar division between “Mekong” and “inland” can be seen in the wet season as well, suggesting that these areas are structurally very different. The inland area is, again, very mixed process-wise, however the first PC describes a process where wCAP and wENV support one another against vulnerability (wUSE): increasing capacity to manage water increases in areas of higher environmental integrity, while the same areas may be more dependent on water for income. The Mekong area is more clear in PC interpretation: wPC1 shows a positive relationship in wCAP and wACC, which directly relates to the socio-economic situation in villages, and wPC2 represents urban/rural divide with increasing wENV affecting negatively on wCAP and wACC. This village location-based process have been identified in the north as well. In the “Mekong” area, wPC3 represents the subsistence/commercial farming divide which has also been identified in the north and in the global PCA.

Finally, in the south, the dominant dry season process (dPC1) is where dRES and dCAP correlate positively. This appears to describe the Bolaven Plateau, which is an important agricultural area. The second and third PC counter-balance: in dPC2, dUSE and dENV contradict, while in dPC3 they are supporting. This is most likely due to interaction between the indicators that make up ENV and USE, with possible reasons being the subsistence/commercial farming or urban/rural divide. For the southern wet season, the wPC1 and wPC2 are broken into small areas with mixed components with the only general process that may be indentified is the area in the south away from Mekong, where wENV and wCAP showing positive relationship. The third process, wPC3, however, shows a uniform process in the entire southern Laos where wUSE and wENV increase as wRES decreases. This may be interpreted as describing a situation where more rural villages engaging in subsistence agriculture are located in areas receiving more rain coupled with shorter dry spells.

The described processes can be summarized by computing local weights using the local component loads and variances from GWPCA. The weights summarize how strongly each WPI component is “present” in the retained PCs. There is a strong agreement between the weights shown in *Figure 5* and the lead components (plots shown in Supplement B). Dry season differences in WPI in northern Laos can largely be explained by CAP, USE and ENV. These three components correlate well with the three main themes of processes identified: poverty (socio-economy, CAP), rurality or degree of development (income vulnerability, USE) and location (environmental integrity and disasters, ENV). RES receives near-zero weight in the North, consistent with the observation that it is irrelevant to the WPI processes in the area. WPI differences in the central-north and the capital region appear to be explained by dUSE and dENV and dACC. This is similar to the north with the exception that here emphasis is put on presence of infrastructure rather than capacity level of villages. In the wet season wRES and wACC are the most important (highest weighted) components, highlighting the importance of access to infrastructure. In addition, also wUSE and wENV are relevant in different locations of the central-north and capital region. Central-south and south are described by the same components apart from the southernmost tip of the country: dRES, dUSE and dENV in the dry, and wCAP, wUSE and wENV in the wet season. Here, as a source of differences between villages, dRES highlights the importance of the amount of water in the dry season, and wCAP the importance of location or rurality (road access).

In addition to the weights, *Figure 5* also presents a local WPI computed from the local weights. An interesting result is that the general structure of higher and lower WPI areas in the global PCA WPI are preserved in the local WPI. The preservation of the structure has also been found by Harris et al (2011) for simulated data. The univariate distribution of WPI computed from the five different weighting schemes are shown in *Figure 6*. An interesting property identified from the figure is that the distribution of WPI weighted by different analytical methods are of the same shape, but whose ranges differ. A major advantage of the analytical weighting is that it highlights differences between villages by resulting in a wider range of WPI values – and in this sense, local WPI creates the widest distribution.

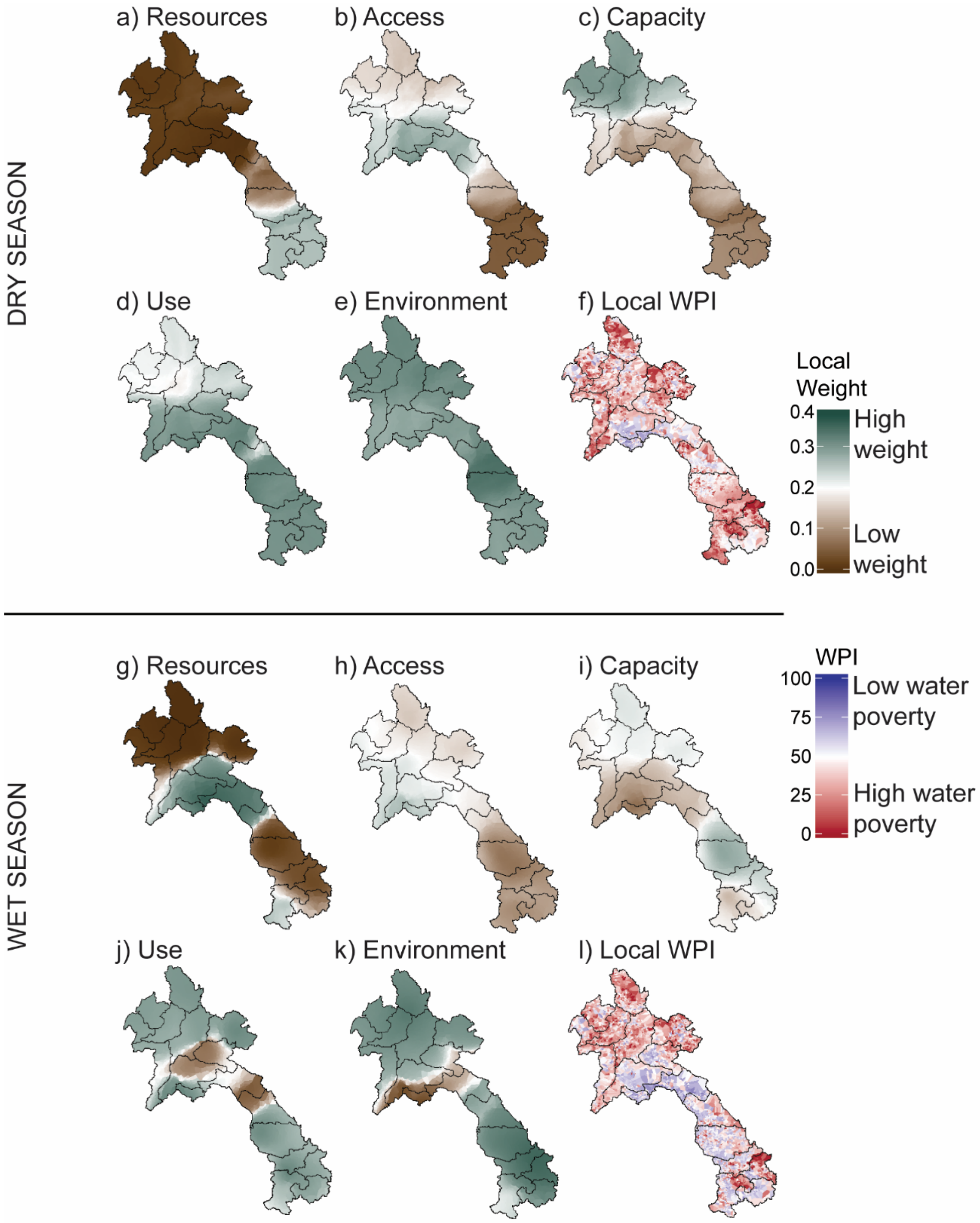

Figure 5. Local WPI component weights for dry (a to e) and wet (g to k) seasons. Locally weighted WPI is shown in plot f) for dry season and plot l) for wet season. Weights below 0.2 (equal weights) are shown in brown and weights above that threshold are shown in green. For the colour reference, see the online version.

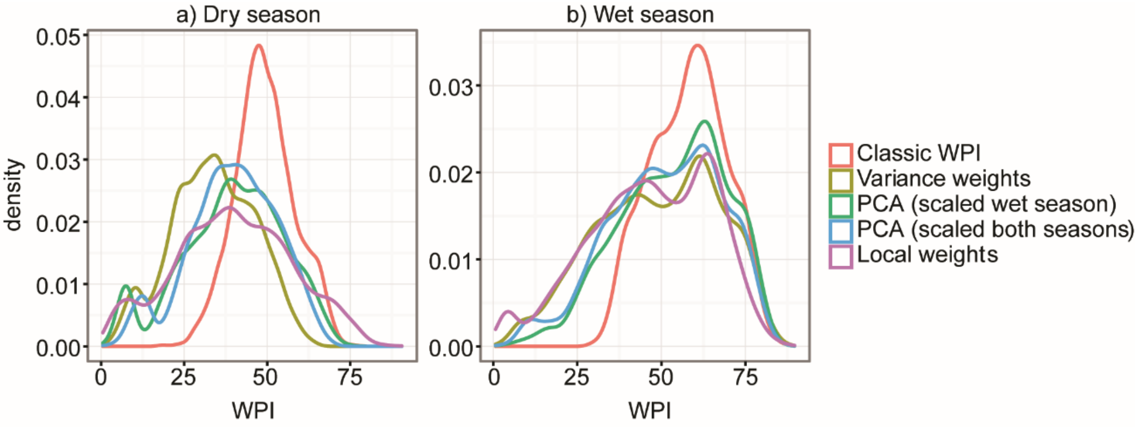

Figure 6. WPI density curves of a) dry and b) wet season. WPI computed by classic additive equal weights, variance based weights, PCA derived weights (from scaled single and both season data) and local weights. For the reference to colour, see the online version.

5 Discussion

No other studies have been conducted in Laos using WPI as a tool, apart from the international comparison (Lawrence et al. 2002). The comparison study, however, was conducted on a global scale, while this study focuses on village-level water poverty with a very different set of indicators. The two other WPI studies made in the Lower Mekong Basin (Guppy 2014; Ty et al. 2010) agree in that rural mountainous villages are more water poor than ones located in fertile lowlands. While WPI cannot be extended outside of the data sources, it is likely that similar processes govern water poverty also in undeveloped neighbouring countries of Cambodia, Myanmar and rural Vietnam. For a more informed discussion, a regional analysis of water poverty with data from the neighbouring countries should be conducted.

Laos is a mostly rural country with a low urbanization rate, and hosts only a few significant sized urban centers, mostly provincial capitals, in addition to the capital city Vientiane. Therefore, no significant differences were identified among the urban centers: the areas around them generally engage in commercial farming, have high WPI and their socio-economic character is comparably good. Rural areas are much more heterogenous. Bolaven Plateau and southern Xayabury districts have very high WPI's. In other areas, proximity to urban centers or location along major roads correspond to higher WPI, which is may be due to better access to markets and lower poverty levels. These villages have also higher development status than more remote villages.

However, the index values derived in this study are largely based on census data which are six to twelve years old, and it has been shown in other research that there has been significant progress in poverty levels since the data was collected (Coulombe et al. 2016; Najdov and Phimmahasay 2016). In addition, due to limitations in data, CAP is also biased to lower scores in the wet season, because river transport could not be considered. It is known to be an important replacement for road access. In future research, newer and alternative data sources should be used to allow for a more thorough temporal assessment than is possible simply over the seasons. The village-level data should additionally be coupled to province and national level data.

Regarding PCA and PCA-based weighting, it was found that they provide a useful data-driven means of highlighting sources of variation. The PCs, in addition to weighting, uncover interesting patterns that drive water poverty, which were tentatively interpreted here as processes. GWPCA extends the approach further by allowing inspection of local differences in these processes rather than dataset-wide trends found using standard PCA. Local weights derived from the local variances and component coefficients seem to be a promising exploratory tool in interpreting the large amount of data produced by GWPCA, and may be a useful addition to the visualization methods (see Section 2.2) suggested by the developers of the method. Translated into practice, local processes identified using GWPCA potentially allow targeting policies in certain regions to address specific processes that drive water poverty.

The results of GWPCA analysis also reveal that, despite being locally weighted, the WPI ranks are comparable to those of global analysis. It is identified as a future research topic to assess whether locally weighted WPI could be statistically sound alternative, despite weights changing for each observation, based on location.

6 Conclusion

The Water Poverty Index was computed for 8215 villages in Laos, and it was found, consistent with expectations, that water poverty is largely a management issue in Laos: resource availability generally is not a problem and the largest cause of variation is in the socio-economic components ACC, CAP and USE. The lowest scoring components (ACC and USE) are the ones where improvement is most important. Spatially, the area along the Mekong river (and especially the capital region) on the Thai border is less water poor, while the mountainous areas on the Vietnamese and Chinese borders are more water poor (*Figure 3*). Temporal differences were also identified, with most important feature being the observation that in villages with low dWPI, increase in wWPI is also low. Process-wise, three main themes driving water poverty in Laos were identified throughout the analysis: socio-economic status and capacity of management, purpose of agriculture (commercial/subsistence), and village location (urban/rural, disaster occurrence).

If PCA is to be used in weighting, it is recommended to perform it on mean-variance scaled components to facilitate examining processes that drive water poverty – differences in variance of unscaled components otherwise mask the processes to some extent. GWPCA can be further used to investigate local processes, giving more detailed information on the spatial differences, as a first step towards helping policy makers to target specific causes of water poverty in specific regions.

REFERENCES

Babel, M., & Wahid, S. (2009). *Freshwater Under Threat: Southeast Asia. Vulnerability Assessment of Freshwater Resources to Environmental Change. Mekong River Basin*. Nairobi, Kenya.

Beilfuss, R., & Triet, T. (2014). *Climate change and hydropower in the Mekong River Basin: a synthesis of research*. https://www.giz.de/en/downloads/giz2014-en-study-climate-change-hydropower-mekong.pdf

Brunsdon, C., Fotheringham, A. S., & Charlton, M. (2002). Geographically weighted summary statistics - a framework for localised exploratory data analysis. *Computers, Environment and Urban Systems*, *26*(6). doi:10.1016/S0198-9715(01)00009-6

Charlton, M., Brunsdon, C., Demšar, U., Harris, P., & Fotheringham, S. (2010). Principal Components Analysis: from Global to Local. *13th AGILE International Conference on Geographic Information Science*, 1–10.

Chatfield, C., & Collins, A. J. (1980). *Introduction to multivariate analysis*. London: Chapman and Hall.

Cho, D. I., Ogwang, T., & Opio, C. (2010). Simplifying the water poverty index. *Social Indicators Research*, *97*(2), 257–267. doi:10.1007/s11205-009-9501-2

Coulombe, H., Epprecht, M., Pimhidzai, O., & Sisoulath, V. (2016). *Where are the poor? Lao PDR 2015 census-based poverty map : province and district level results*. Washington, D.C. http://documents.worldbank.org/curated/en/477381468415961977/Where-are-the-poor-Lao-PDR-2015-census-based-poverty-map-province-and-district-level-results

Darby, S. E., Hackney, C. R., Leyland, J., Kummu, M., Lauri, H., Parsons, D. R., et al. (2016). Fluvial sediment supply to a mega-delta reduced by shifting tropical-cyclone activity. *Nature*, *539*(7628), 276–279. doi:10.1038/nature19809

Demšar, U., Harris, P., Brunsdon, C., Fotheringham, A. S., & McLoone, S. (2013). Principal

Component Analysis on Spatial Data: An Overview. *Annals of the Association of American Geographers*, *103*(1). doi:10.1080/00045608.2012.689236

Falkenmark, M., Lundqvist, J., & Widstrand, C. (1989). Macro-scale water scarcity requires micro-scale approaches. *Natural Resources Forum*, *13*(4), 258–267. doi:10.1111/j.1477-8947.1989.tb00348.x

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. New York: Wiley.

Garriga, R. G., & Foguet, A. P. (2010). Improved Method to Calculate a Water Poverty Index at Local Scale. *Journal of Environmental Engineering*, *136*, 1287–1298. doi:10.1061/(ASCE)EE.1943-7870.0000255

Getis, A. (2010). Spatial Autocorrelation. In M. Fischer & A. Getis (Eds.), *Handbook of Applied Spatial Analysis* (pp. 255–278). Berlin: Springer Berlin Heidelberg.

Guppy, L. (2014). The Water Poverty Index in rural Cambodia and Viet Nam: A holistic snapshot to improve water management planning. *Natural Resources Forum*, *38*(3). doi:10.1111/1477-8947.12051

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis* (Sixth.). Upper Saddle River, New Jersey: Pearson Prentice Hall.

Hajkowicz, S. (2006). Multi-attributed environmental index construction. *Ecological Economics*, *57*(1), 122–139. doi:10.1016/j.ecolecon.2005.03.023

Harris, P., Brunsdon, C., & Charlton, M. (2011). Geographically weighted principal components analysis. *International Journal of Geographical Information Science*, *25*(10), 1717–1736. doi:10.1080/13658816.2011.554838

Harris, P., Clarke, A., Juggins, S., Brunsdon, C., & Charlton, M. (2014). Geographically weighted methods and their use in network re-designs for environmental monitoring. *Stochastic Environmental Research and Risk Assessment*, *28*(7), 1869–1887. doi:10.1007/s00477-014-0851-1

Harris, P., Clarke, A., Juggins, S., Brunsdon, C., & Charlton, M. (2015). Enhancements to a geographically weighted principal component analysis in the context of an application to an environmental data set. *Geographical Analysis*, *47*(2), 146–172. doi:10.1111/gean.12048

Heidecke, C. (2006). Development and evaluation of a regional water poverty index for Benin, (January), 1–55.

Jemmali, H., & Matoussi, M. S. (2013). A multidimensional analysis of water poverty at local scale: Application of improved water poverty index for Tunisia. *Water Policy*, *15*(1), 98–115. doi:10.2166/wp.2012.043

Jemmali, H., & Sullivan, C. A. (2014). Multidimensional Analysis of Water Poverty in MENA Region: An Empirical Comparison with Physical Indicators. *Social Indicators Research*, *115*(1), 253–277. doi:10.1007/s11205-012-0218-2

Jolliffe, I. T. (2002). Principal Component Analysis, Second Edition. *Encyclopedia of Statistics in Behavioral Science*, *30*(3), 487. doi:10.2307/1270093

Komnenic, V., Ahlers, R., & Zaag, P. van der. (2009). Assessing the usefulness of the water poverty index by applying it to a special case: Can one be water poor with high levels of access? *Physics and Chemistry of the Earth*, *34*(4–5), 219–224. doi:10.1016/j.pce.2008.03.005

Koponen, J., Lauri, H., Veijalainen, N., & Sarkkula, J. (2010). *HBV and IWRM Watershed Modelling User Guide*. Phnom Penh.

Lao Statistics Bureau. (2005). Census of Population and Housing 2005. Vientiane, Laos: Lao Statistics Bureau. http://www.decide.la/

Lao Statistics Bureau. (2011). Lao Agriculture Census 2010/2011. Vientiane, Laos: Lao Statistics Bureau. http://www.decide.la/

Lauri, H., Räsänen, T. A., Kummu, M., Lauri, H., Räsänen, T. A., & Kummu, M. (2014). Using Reanalysis and Remotely Sensed Temperature and Precipitation Data for Hydrological Modeling in Monsoon Climate: Mekong River Case Study. *Journal of Hydrometeorology*, *15*(4), 1532–1545. doi:10.1175/JHM-D-13-084.1

Lawrence, P., Meigh, J., & Sullivan, C. (2002). The Water Poverty Index: An International Comparison. *Keele Economics Research Papers*, *19*(October), 17. doi:10.1111/1477-8947.00054

Lloyd, C. D. (2010). Analysing population characteristics using geographically weighted principal components analysis: A case study of Northern Ireland in 2001. *Computers, Environment and Urban Systems*, *34*(5). doi:10.1016/j.compenvurbsys.2010.02.005

Mekong River Commission. (2007). *Diagnostic Study of Water Quality in Lower Mekong Basin. MRC Technical Paper No. 15.* Vientiane, Laos.

Mekong River Commission. (2011). *Planning Atlas of the Lower Mekong River Basin*. Vientiane, Laos.

Molle, F., & Mollinga, P. (2003). Water poverty indicators: Conceptual problems and policy issues. *Water Policy*, *5*(5–6), 529–544.

Najdov, E., & Phimmahasay, K. (2016). *Lao economic monitor - Challenges in promoting more inclusive growth and shared prosperity : thematic section - drivers of poverty reduction in Lao PDR*. Washington, D.C. http://documents.worldbank.org/curated/en/515521468197368035/Lao-economic-monitor-Challenges-in-promoting-more-inclusive-growth-and-shared-prosperity-thematic-section-drivers-of-poverty-reduction-in-Lao-PDR

Pérez-Foguet, A., & Garriga, R. G. (2011). Analyzing Water Poverty in Basins. *Water Resources Management*, *25*(14), 3595–3612. doi:10.1007/s11269-011-9872-4

Räsänen, T. A., Someth, P., Lauri, H., Koponen, J., Sarkkula, J., & Kummu, M. (2017). Observed river discharge changes due to hydropower operations in the Upper Mekong Basin. *Journal of Hydrology*, *545*, 28–41. doi:10.1016/j.jhydrol.2016.12.023

Sullivan, C. (2002). Calculating a Water Poverty Index. *World Development*, *30*(7), 1195–1210. doi:10.1016/S0305-750X(02)00035-9

Sullivan, C., & Meigh, J. (2007). Integration of the biophysical and social sciences using an indicator approach: Addressing water problems at different scales. *Integrated Assessment of Water Resources and Global Change: A North-South Analysis*, (February 2006), 111–128. doi:10.1007/978-1-4020-5591-1-8

Sullivan, C., Meigh, J., Giacomello, M., Fediw, T., Lawrence, P., Samad, M., et al. (2003). The water poverty index: Development and application at the community scale. *Natural Resources Forum*, *27*, 189–199. doi:10.1111/1477-8947.00054

Tang, X., & Feng, Q. (2016). The temporal-spatial assessment of water scarcity with the Water Poverty Index: a study in the middle basin of the Heihe River, northwest China. *Water Science and Technology-Water Supply*, *16*(5), 1266–1276. doi:10.2166/ws.2016.053

The United Nations in Lao PDR. (2015). *Country Analysis Report : Lao PDR*. Vientiane. http://www.la.undp.org/content/lao_pdr/en/home/library/mdg/country-analysis-report.html

Ty, T. Van, Sunada, K., Ichikawa, Y., & Oishi, S. (2010). Evaluation of the state of water resources using Modified Water Poverty Index: a case study in the Srepok River basin, Vietnam – Cambodia. *International Journal of River Basin Management*, *8*(April 2013), 305–317. doi:10.1080/15715124.2010.523004

Wei, C., Cabrera-Barona, P., & Blaschke, T. (2016). Local geographic variation of public services inequality: Does the neighborhood scale matter? *International Journal of Environmental Research and Public Health*, *13*(10). doi:10.3390/ijerph13100981

Wildlife Conservation Society - WCS, & Center for International Earth Science Information Network - CIESIN - Columbia University. (2005). Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Footprint Dataset (Geographic). Palisades, NY: ASA Socioeconomic Data and Applications Center (SEDAC).

Zhang, Q., Liu, B., Zhang, W., Jin, G., & Li, Z. (2015). Assessing the regional spatio-temporal pattern of water stress: A case study in Zhangye City of China. *Physics and Chemistry of the Earth*, *79–82*, 20–28. doi:10.1016/j.pce.2014.10.007

SUPPLEMENT A – Component and indicator processing

Before applying PCA to the components to compute the weights, the applicability of PCA was assessed for each WPI component using the determinant of correlation matrix between indicators, together with Fligner-Killeen test of sphericity. The aim of this step is to ensure that indicators within an individual component measure different things and to avoid double counting (redundant indicators) (Hajkowicz 2006). The Fligner-Killeen test is used instead of the Bartlett's test used by Jemmali and Sullivan (2014) and Cho et al (2010) due to significant non-normality in univariate indicator distributions. Additionally, the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (MSA) was addressed as outlined by Hair et al (2006). MSA indicates whether there are significant intercorrelations among variables to make PCA an appropriate method to use.

Table 7. Correlation and factorability tests for the components of dry season and wet season WPI. RES, ACC, CAP, USE and ENV signify the WPI components Resources, Access, Capacity, Use and Environment, respectively.

	dRES	dACC	dCAP	dUSE	dENV	wRES	wACC	wCAP	wUSE	wENV
Determinant of Correlation Matrix	0.96	0.90	0.35	1.00	0.92	0.98	0.90	0.25	0.99	0.90
MSA	0.48	0.43	0.67	0.50	0.57	0.49	0.43	0.73	0.50	0.57
Fligner-Killeen test										
chi^2	7849.87	574.32	6292.62	7470.99	1550.02	15328.02	574.32	156.84	3556.10	287.16
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Degrees of Freedom	2.00	2.00	3.00	1.00	2.00	2.00	2.00	3.00	1.00	2.00

The results for each component is shown in *Table 7*. The Fligner-Killeen test shows that there are significant non-zero correlations present in each component. However, the determinant of correlation matrix shows that this is problematic only in CAP, with values of 0.35 and 0.25 in dry and wet seasons respectively. As mentioned in Section 2 of the main text, PCA can also be used to derive new, uncorrelated indicators from the original CAP selection. The MSA for CAP is 0.67 for dry and 0.73 for wet season, falling in the acceptable range of 0.50 or above, meaning that we can meaningfully perform PCA on the four selected indicators in CAP (Hair, et al., 2006). For this study, we consistently retain, in each analysis, the number of PCs that explain more than 70% of total variation. Therefore, two PCs (the new uncorrelated variables) were retained, explaining a total of 80.5% of total variance within CAP indicators.

MSA analysis shows that RES and ACC components for both dry and wet seasons are below 0.50 and thus fall in the unacceptable range. Hair et al (2006) suggest that in order to increase MSA, we should eliminate indicators from the component. An alternative is to reject the components RES and ACC entirely, as Jemmali and Sullivan (2014) and Cho et al (2010) did in their analyses. In this study the first alternative is adopted due to RES being the component that varies most between the seasons. Therefore, water availability per capita was removed from RES because of its score distribution – 91.5% of villages score 100 in the dry season and 99.2% in the wet season. This indicates that the quantity of water is an issue only in a limited number of villages, which is consistent with the study by Ty et al (2010) in Cambodia-Vietnam. Presence of irrigation facilities was removed from ACC because USE already includes a variable of irrigation penetration. Elimination of the aforementioned indicators increased MSA to 0.50 for both RES and ACC.

The same analysis of correlations and MSA was also conducted for the aggregated component scores to make sure that the weights can be calculated using PCA. Fligner-Killeen test was significant (p-value < 0.01) and thus there are significant correlations among components of both dWPI and wWPI with both determinants of correlation matrices being 0.68. MSA for both fall in the acceptable range, being 0.535 for dry and 0.524 for wet season. This confirms that PCA can be meaningfully performed on WPI components to compute the weights. *Table 8* shows the correlation matrix of the dry and wet season WPI components. Of these components, only ACC and CAP show significant correlation, with the rest being below 0.2 level.

Table 8 Correlation matrix between components for dry (d-) and wet (w-) season. RES, ACC, CAP, USE and ENV signify the WPI components Resources, Access, Capacity, Use and Environment, respectively.

	dRES	**dACC**	**dCAP**	**dUSE**	**dENV**
dryRES	1.000	-0.129	-0.164	-0.158	0.097
dryACC	-0.129	1.000	0.472	0.114	-0.041
dryCAP	-0.164	0.472	1.000	0.088	-0.012
dryUSE	-0.158	0.114	0.088	1.000	-0.045
dryENV	0.097	-0.041	-0.012	-0.045	1.000
	wRES	**wACC**	**wCAP**	**wUSE**	**wENV**
wetRES	1.000	-0.042	-0.001	-0.004	-0.049
wetACC	-0.042	1.000	0.480	0.166	-0.059
wetCAP	-0.001	0.480	1.000	0.122	-0.002
wetUSE	-0.004	0.166	0.122	1.000	-0.042
wetENV	-0.049	-0.059	-0.002	-0.042	1.000

SUPPLEMENT B – GWPCA figures

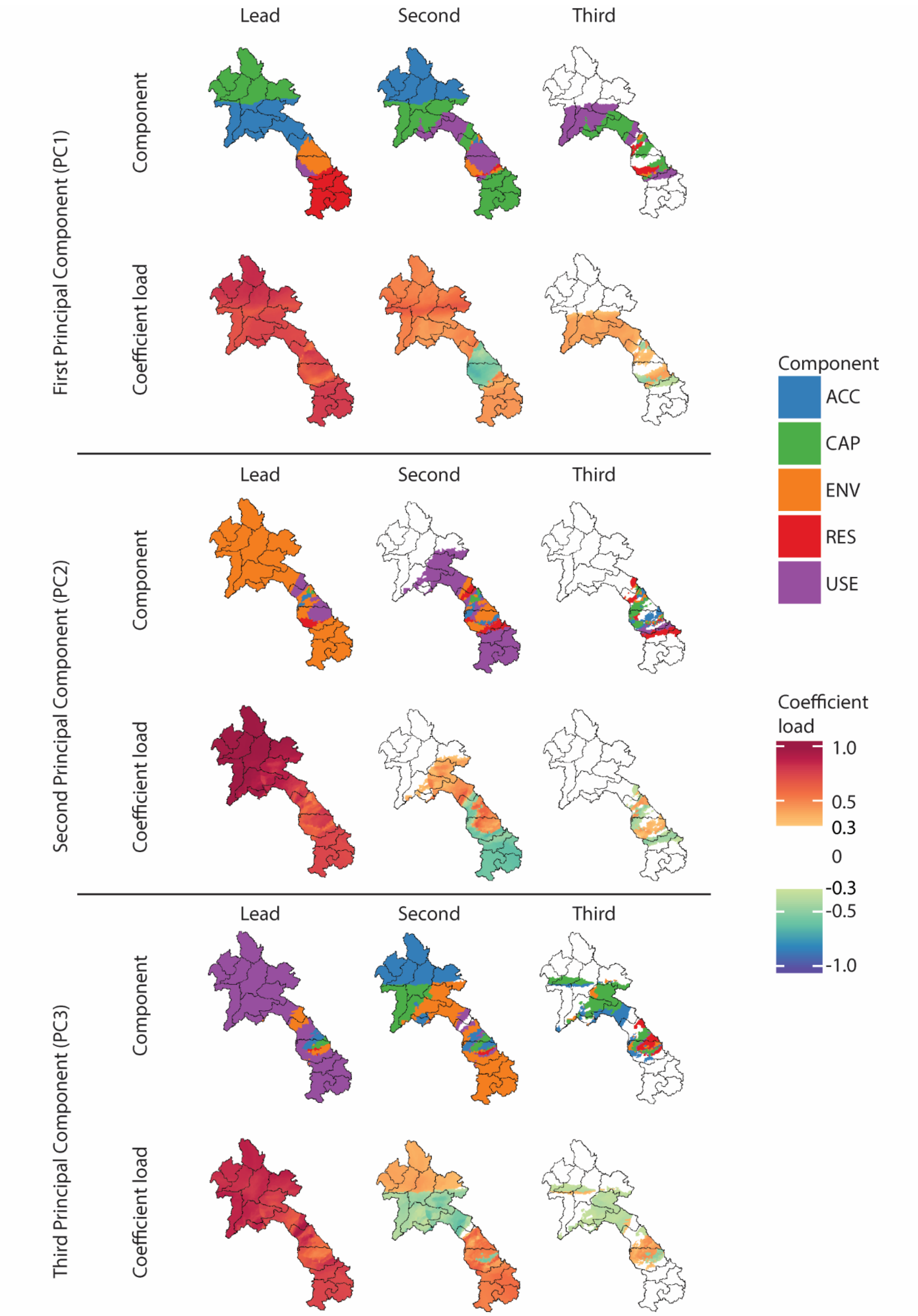

Figure 7. The first three highest loaded components in the retained local principal components and their loading coefficients for dry season scaled GWPCA. Items with a load smaller than 0.3 are excluded and shown in white.

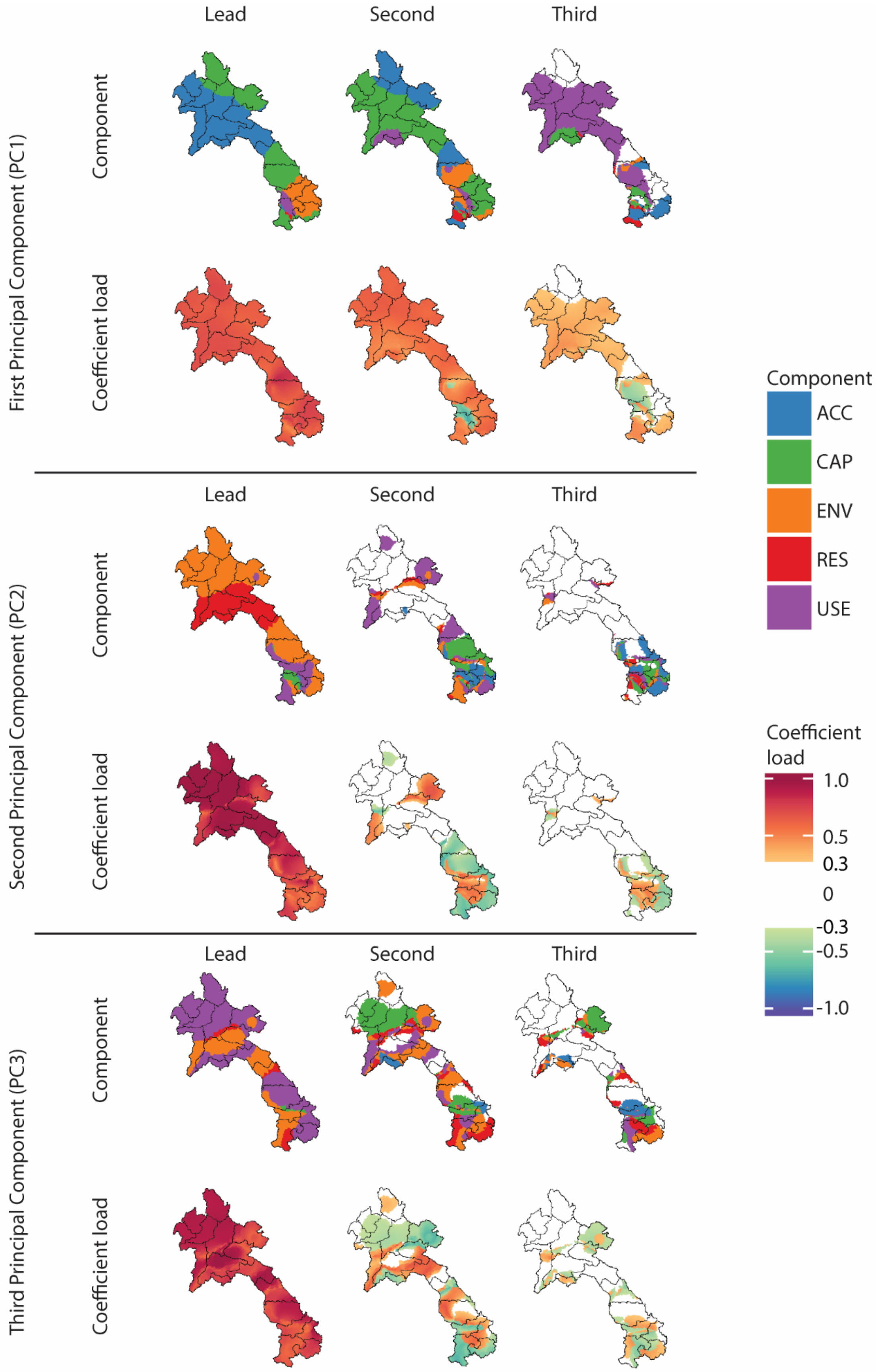

Figure 8. The first three highest loaded components in the retained local principal components and their loading coefficients for wet season scaled GWPCA. Items with a load smaller than 0.3 are excluded and shown in white.