Position Preserving Multi-Output Prediction

Zubin Abraham[1], Pang-Ning Tan[1], Perdinan[1], Julie Winkler[1], Shiyuan Zhong[1], and Malgorzata Liszewska[2]

[1] Michigan State University, USA
{abraha84,perdinan,winkler,zhongs}@msu.edu, ptan@cse.msu.edu
[2] University of Warsaw, Poland
m.liszewska@icm.edu.pl

Abstract. There is a growing demand for multiple output prediction methods capable of both minimizing residual errors and capturing the joint distribution of the response variables in a realistic and consistent fashion. Unfortunately, current methods are designed to optimize one of the two criteria, but not both. This paper presents a framework for multiple output regression that preserves the relationships among the response variables (including possible non-linear associations) while minimizing the residual errors of prediction by coupling regression methods with geometric quantile mapping. We demonstrate the effectiveness of the framework in modeling daily temperature and precipitation for climate stations in the Great Lakes region. We showed that, in all climate stations evaluated, the proposed framework achieves low residual errors comparable to standard regression methods while preserving the joint distribution of the response variables.

1 Introduction

Multiple output regression (MOR) is the task of inferring the joint values of multiple response variables from a set of common predictor variables. The response variables are often related, though their true relationships are generally unknown *a priori*. An example application of multiple output regression is to simultaneously estimate the projected future values of temperature, precipitation, and other climate variables needed for climate change impact, adaptation and vulnerability (CCIAV) assessments. The projected values are used as the driving input variables for phenological and hydrological models to simulate the responses of the ecological system to future climate change scenarios. To ensure the projected values are realistic, there are certain constraints on the relationship among the response variables that must be preserved; e.g., minimum temperature must not exceed maximum temperature or liquid precipitation should be zero when temperature is below freezing. While there have been numerous multiple output regression methods developed in recent years [7,20,4,18,12], most of them are focused on fitting the conditional mean or preserving covariance structure of the outputs. Such methods do not adequately capture the full range of variability in the joint output distribution, as illustrated in Figure 1(a).

The inability of standard regression-based approaches to reproduce the shape of the true distribution of output variables, even for univariate response

H. Blockeel et al. (Eds.): ECML PKDD 2013, Part II, LNAI 8189, pp. 320–335, 2013.

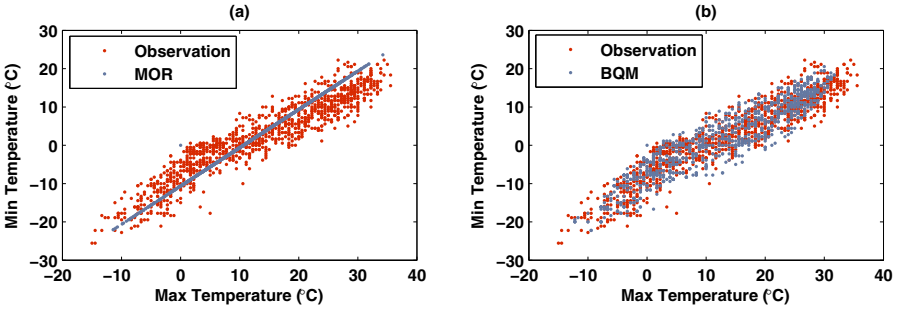

Fig. 1. Scatter plot of observed daily maximum and minimum temperature at a climate station in Michigan, USA

variables, is well-documented [2]. Univariate *distribution-driven approaches* such as quantile mapping (QM) [13] and statistical asynchronous regression (SAR) [17] have been developed to address this limitation, but the accuracy of these approaches is generally poor since they are not designed to minimize residual errors. Quantile mapping approaches map a univariate predictor variable x to its corresponding response variable y by transforming the cumulative distribution function (CDF) of x to match that of y. More recently, a bivariate quantile mapping approach (BQM) (see Figure 1(b)) has been developed to generate bivariate response values that mimic the joint distribution of the observed response data [11]. However, as will be shown in this paper, the residual error is significantly worse when compared to regression-based methods because the position and rank correlation between the predictor and response variables remain invariant under QM-based transformation, which in turn, hinders its ability to minimize residual errors. Thus, unless the predictor variable has a high rank correlation with the response variable, the residual error upon applying QM-based approaches is likely to be large.

This suggests a possible hybrid approach to improve both the residual errors and distribution fitting is by first applying a regression-based method to transform the predictor variables so that their rank correlation with respect to the response variable is high, before applying quantile mapping to adjust for the fit in distribution. However, maximizing the rank correlation of the data points is necessary but not sufficient condition for improvement in the residuals for QM, unless the response values of the data points are uniformly spaced. Hence, the need for position regularization, that would prioritize the prediction accuracy of data points whose position, when incorrectly estimated, results in high residual. The term 'position' here refers to the geometric quantile of a data point with respect to a multivariate distribution, which is analogous to the quantile of a data point in the case of univariate distribution. In this paper, we present a position-regularized, multi-output prediction framework called Multi-Output Contour Regression (MCR), that addresses the dual objective of preserving the associations among the multiple output variables as well as minimizing

residuals. MCR is able to achieve the dual objective by applying a novel, position-regularized regression method, followed by geometric quantile mapping (GQM) to improve the fit in distribution. The position-regularized regression helps to alleviate the limitation associated with the rank invariant property of QM, which contributes to the high residuals of QM-based approaches. MCR additionally addresses the challenge of ensuring that its prediction of the response variables will always abide by the constraints of the actual response data. MCR is also not limited by the number of predictor variables that may be used nor does it require them to have high correlation with the response variables, unlike quantile mapping. The flexible nature of our framework allows for the incorporation of other loss functions such as the L_1 loss used in quantile regression[1].

2 Related Work

Supervised learning methods for predicting continuous-valued outputs may be categorized as either *accuracy-driven* or *distribution-driven*. Accuracy-driven approaches such as multiple linear regression (MLR), lasso regression, neural networks, and analog methods [13] are commonly used with emphasis on minimizing sum-square residual (SSR) errors. In contrast, distribution-driven approaches focus on reproducing the distribution characteristics of the output variable. These approaches include quantile mapping (QM) [13], Equidistant CDF Matching (EDCDFm), statistical asynchronous regression (SAR) [17] and the transfer functions proposed by Piani et al. [19]. These approaches are applicable even when the predictor and response variables are asynchronous and are generally susceptible to high residual errors. Given the drawbacks of accuracy-driven and distribution-driven approaches, a hybrid method known as Contour Regression (CR) [2] was developed to simultaneously minimize error and preserve the shape of the fitted distribution. CR extends the loss functions of standard regression methods (including linear and quantile regression) to regularize the area between the CDF of the response variable and the CDF of the predicted output.

In addition to the single output regression (SOR) approaches, techniques for inferring multiple response variables (MOR) simultaneously have been developed, including multi-output regression [10] and structured output regression [5]. A number of these techniques focus on penalizing the regression coefficients using low rank methods such as reduced rank regression [12]. However, these approaches do not consider the correlation among the output variables. Another common approach to multiple output prediction is to penalize the shared input space, for co-linearity, such as partial least square regression discriminant analysis (PLSDA) [18]. However these models, too, do not capture the association among response variables. Curds and Whey is an example of regression based approach that considers the output correlation [7]. However, it assumes the relationship among the response variables is linear. Multiple output SVR is another approach that takes advantage of correlation among response variables and extends Support Vector Regression (SVR) to multi-output systems by employing

[1] We omit the derivation for other loss function in this paper due to lack of space.

co-kriging, to account for the cross covariances between different response variables [20]. Group lasso [14], LL-MIMO [6], gaussian process MOR [3] are other examples of MOR.

However, none of the these approaches preserve the full range of variability of the joint distribution of the response variables. He et al. [11] proposed bivariate quantile mapping to extend QM to bivariate space. The method uses the intuition proposed by Buja et al. [8] regarding geometric quantiles. While this approach is capable of capturing the distribution characteristics of bivariate response variables, similar to QM, it is susceptible to high residual errors.

3 Preliminaries

Let $\mathbf{X} = [\mathbf{x}_1, .., \mathbf{x}_n]^T$ be an $(n \times d)$ data matrix and $\mathbf{Y} = [\mathbf{y}_1, .., \mathbf{y}_n]^T$ be the corresponding $(n \times q)$ response matrix, such that $\mathbf{x}_i \in \Re^d$ and $\mathbf{y}_i \in \Re^q$ are column vectors representing the respective values of predictor and response variables for the i^{th} data point. The objective of multi-output regression (MOR) is to learn a target function $h(\mathbf{x}, \Omega)$ that best estimates the multi-output response $\mathbf{y}$, where $\Omega = (\omega_1, .., \omega_q)$ is the parameter set of the target function.

For a univariate random variable $X \in \Re$, let $F_X(x)$ be its cumulative distribution function (CDF), i.e., $F_X(x) = P(X \leq x)$. The corresponding α-quantile of X is given by $\inf\{x \in \Re : F_X(x) \geq \alpha\}$. Intuitively, each quantile indicates the value in which a certain fraction of the data points are below it, and thus, provides a measure of its position in the data. For example, the median, which is equivalent to the 0.5-quantile, is the central location of the distribution. More generally, the position [16] of data point $\mathbf{z}$ relative to a set of points $\mathbf{Z} = (\mathbf{z}_1, .., \mathbf{z}_m)^T$ is given by

$$\mathbf{p}_{\mathbf{Z}}(\mathbf{z}) = \tfrac{1}{m} \textstyle\sum_{i=1}^{m} \eta(\mathbf{z} - \mathbf{z}_i) \qquad \text{where} \quad \eta(\mathbf{w}) = \begin{cases} \frac{\mathbf{w}}{\|\mathbf{w}\|}, & \text{if } \mathbf{w} \neq \mathbf{0} \\ 0, & \text{if } \mathbf{w} = \mathbf{0} \end{cases}$$

For univariate data, the position $p_Z(z)$ is equal to $2F_Z(z) - 1$, where $F_Z(z)$ is the cumulative distribution function of Z. The multi-dimensional equivalent of quantile function is geometric quantile [9].

Distribution correction methods such as quantile mapping is only applicable if one can match the position of a data point in one univariate distribution (say for x) to its corresponding position in another univariate distribution (say for y). This is possible using the preceding definition of position for univariate data since the values of p_Z are always fixed in the range between $[-1, +1]$ irrespective of the values in Z. Unfortunately, when extended to multivariate positions, the range of values for p_Z may vary depending on the values in Z. To overcome this problem, He et al. [11] introduce the notion of a stationary position by iteratively applying the following position transformation function until convergence:

$$\mathbf{p}_Y^k(\mathbf{z}) = \frac{1}{\kappa n} \sum_{i=1}^{n} \frac{\mathbf{p}_Y^{k-1}(\mathbf{z}) - \mathbf{p}_Y^{k-1}(\mathbf{y}_i)}{\| \mathbf{p}_Y^{k-1}(\mathbf{z}) - \mathbf{p}_Y^{k-1}(\mathbf{y}_i) \|}, \qquad \mathbf{p}_Y^1(\mathbf{z}) = \frac{1}{\kappa n} \sum_{i=1}^{n} \frac{\mathbf{z} - \mathbf{y}_i}{\| \mathbf{z} - \mathbf{y_i} \|} \quad (1)$$

Here each component in $\mathbf{y}_i$ must be converted to its marginal rank first before applying the position transformation function. Marginal rank refers to the rank of the data point divided by the largest rank and then normalized to the range $[-1, 1]$. The normalization is done to negate the effect of variables having values that correspond to different ranges. Data points with normalized marginal rank close to ± 1 correspond to extreme values for the particular variable, while those close to $\mathbf{0}$ are located near the median of the distribution. In practice, the number of iterations needed to reach a stationary distribution is quite small, typically $K > 5$ [11]. For univariate data, it can be shown that $\mathbf{P}^k$ reaches a stationary distribution at $k = 1$.

The term κ in Equation (1) is a normalization factor to ensure the distribution of the geometric positions is supported in a q-dimensional unit hypersphere. In the case of bivariate response variable $\mathbf{Y}$, the stationary geometric quantile distribution is circularly symmetric around the origin, with the radial density of $r/\sqrt{1-r^2}$ for $r \in (0, 1)$ [11]. Therefore,

$$\kappa = \int_0^1 \frac{r}{\sqrt{1-r^2}} dr \Rightarrow \kappa = \frac{\pi}{4}$$

In this paper, we denote the position of the multivariate data points in $\mathbf{Y}$ as $\mathbf{P}_Y = [\mathbf{p}_Y(\mathbf{y}_1), .., \mathbf{p}_Y(\mathbf{y}_n)]^T$, where $\mathbf{p}_Y(\mathbf{y}_i) \in [-1, 1]^q$. We also use the notation $\mathbf{z}_{XY} = \mathbf{p}_X^{-1}(\mathbf{p}_Y(\mathbf{y}))$ to represent a point in the domain of $\mathbf{X}$ that has the same geometric quantile position as the data point $\mathbf{y}$ in $\mathbf{Y}$, i.e., $\mathbf{p}_X(\mathbf{z}_{XY}) = \mathbf{p}_Y(\mathbf{y})$. Consequently, $\mathbf{z}_{YY}(\mathbf{y}_i) = \mathbf{y}_i$. Finally, let $\mathbf{Z}_{XY} = [\mathbf{z}_{XY}(\mathbf{y}_1)^T, .., \mathbf{z}_{XY}(\mathbf{y}_n)^T]^T$ be the geometric quantiles in X that correspond to the data points in Y.

3.1 Quantile Mapping-Based Approaches

Quantile mapping transforms a univariate predictor variable X to its corresponding response variable Y by adjusting the cumulative distribution function F_X to match that of F_Y:

$$QM : \hat{y} = F_Y^{-1}(F_X(x)) \tag{2}$$

It can be shown that QM preserves the rank correlation[2] between the variables. For instance, consider the example in Table 1 where $\mathbf{y}$ is the response variable and $\mathbf{x}_1$, $\mathbf{x}_2$ are two independent predictor variables. Let QM($\mathbf{x}_1$) and QM($\mathbf{x}_2$) be the corresponding QM outputs for $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively. If we sort the vectors in ascending order, it is easy to see that the resulting rank vectors are invariant under QM transformation. As a result, the rank correlation between $\mathbf{x}_1$ (or $\mathbf{x}_2$) and $\mathbf{y}$ is identical to the rank correlation between QM($\mathbf{x}_1$) (or QM($\mathbf{x}_2$)) and $\mathbf{y}$. Furthermore, the empirical CDF for QM($\mathbf{x}_1$) as well as QM($\mathbf{x}_2$) are identical to that for $\mathbf{y}$, i.e., $F_Y = F_{\mathrm{QM}(\mathbf{x}_1)} = F_{\mathrm{QM}(\mathbf{x}_2)}$.

[2] Examples of rank correlation measures include Kendall τ and Spearman's ρ coefficients.

Even though quantile mapping was able to replicate the empirical distribution of $\mathbf{y}$ perfectly, QM($\mathbf{x}_1$) has a higher residual error than QM($\mathbf{x}_2$). This can be explained by the lower rank correlation between $\mathbf{x}_1$ and $\mathbf{y}$ compared to the rank correlation between $\mathbf{x}_2$ and $\mathbf{y}$. Note that the inverse relationship between rank correlation and residual error holds only if the values of the response variable are uniformly spaced. For example, if the response value y for the fourth data point changes from 0.4 to 0.7, the residual error for QM($\mathbf{x}_2$) increases from 0.02 to 0.32, and is larger than the residual error for QM($\mathbf{x}_1$), which remains at 0.06. In this case, a high rank correlation for $\mathbf{x}_2$ does not translate to lower residual error when applying quantile mapping. A formal proof showing the relationship between rank correlation and residual error for uniformly spaced data is given in the next section.

Table 1. Quantile Mapping

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{y}$	QM($\mathbf{x}_1$)	QM($\mathbf{x}_2$)
0.6	0.7	0.2	0.1	0.2
0.8	0.6	0.1	0.3	0.1
0.7	0.9	0.3	0.2	0.4
0.9	0.8	0.4	0.4	0.3
		SSR=	0.06	0.02

Table 2. Quantile Mapping

$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{y}$	QM($\mathbf{x}_3$)	QM($\mathbf{x}_4$)
0.7	0.6	0.2	0.2	0.1
0.6	0.7	0.1	0.1	0.2
0.9	0.8	0.3	0.7	0.3
0.8	0.9	0.7	0.3	0.7
		SSR=	0.32	0.02

Since most data sets are non-uniform, maximizing rank correlation is not a sufficient condition to ensure a low residual error. Nevertheless, we observe that data points associated with quantiles that are located in sparse regions (i.e., far from their next closest quantiles) will contribute to higher residual error when incorrectly ranked compared to data points associated with quantiles located in dense regions. This is demonstrated by the example shown in Table 2, where both $\mathbf{x}_3$ and $\mathbf{x}_4$ have the same rank correlation with respect to the response variable $\mathbf{y}$, yet have different SSR. The response values for the first three data points (0.2, 0.1, and 0.3) are closer to each other than the last data point (0.7). An incorrect ranking of the fourth data point will lead to much higher residual error compared to the first three data points. Since $\mathbf{x}_3$ ranked the fourth data point incorrectly, its residual error is larger than $\mathbf{x}_4$ even though they both have the same rank correlation. This suggests a possible heuristic for improving both rank correlation and residual error by emphasizing on data points that contribute to high residual errors in prediction if ranked incorrectly.

3.2 Rank Correlation and Residual Errors of Quantile Mapping

This section presents several properties of the QM approach with respect to the rank correlation and residual error of its output. First, we show that quantile mapping preserves the rank correlation between the predictor and response variables.

Proposition 1. *Rank correlation is invariant under QM transformation if the values of the predictor and response variables in a data set are unique.*

Proof. Consider a data set $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$ that contains n points. Let $\hat{y}_i$ be the quantile mapped value for the data point with predictor variable x_i. To prove that rank correlation is invariant under QM transformation, it is sufficient to show that the rank for x_i is identical to the rank of $\hat{y}_i$ after quantile mapping. Without loss of generality, assume the data points in $\mathcal{D}$ are sorted in increasing order of their x values. Thus, the rank for data point x_i is i (since the x values are unique). Equation (2) can be rewritten as follows

$$F_Y(\hat{y}_i) = F_X(x_i)$$

Since $F_X(x_i) = i/n$, therefore $F_Y(\hat{y}_i) = F_X(x_i) = i/n$. Given that the response values y_i are distinct, the rank for $\hat{y}_i$ is also i. ♢

Next, we illustrate the relationship between rank correlation and residual error of QM output for data sets with uniformly spaced response values.

Proposition 2. *The SSR of QM output is negatively proportional to the rank correlation of the input and a uniformly spaced response data.*

Proof. Given n data points, let r_i and s_i be the respective ranks of the i^{th} input data point x_i and the corresponding response output y_i. Without loss of generality, we assume that each data point has a unique rank. Since $\mathbf{y}$ is uniformly spaced, $y_i = s_i c_1 + c_0$, where c_0 and c_1 are constants. Similarly, the QM output $\hat{y}_i = r_i c_1 + c_0$. The Spearman rank correlation can be written

$$\rho = \frac{\sum_i (r_i - \bar{r})(s_i - \bar{s})}{\sqrt{\sum_i (r_i - \bar{r})^2 \sum_i (s_i - \bar{s})^2}}$$

We have $\rho = (1/c_2)(\sum_i r_i s_i + c_3)$, where, c_2 and c_3 are constant for a fixed n. Given, $SSR = \sum_i (y_i - \hat{y}_i)^2$ and the QM output $\hat{\mathbf{y}}$ is a reordered instance of $\mathbf{y}$, we have $SSR = 2(\sum_i y_i^2 - \sum_i y_i \hat{y}_i)$. $\sum_i y_i \hat{y}_i = (c_1^2) \sum_i r_i s_i + c4$. where, c_4 is a constants for a fixed n. Therefore, $SSR = 2(\sum_i y_i^2 - (c_1^2 c_2)\rho - c_3 - c_4)$. Since, c_2 is a positive constant, $(c_1^2 c_2)$ will always be positive. Hence, SSR is negatively proportional to ρ when $\mathbf{y}$ is uniformly spaced. ♢

We next show that the output of QM that perfectly replicated the response variable can be improved to have lower residual errors by correcting the ranks of the predictor variable to better match the response variable.

Proposition 3. *Correcting the ranks of data points in* $\mathbf{x}$ *that do not match the rank of the corresponding data point in* $\mathbf{y}$*, maintains, if not, improves the SSR of QM output.*

Proof. Let elements of $R_{\mathbf{x1}}$, $R_{\mathbf{x2}}$ and O be the quantile positions of data points in **x1**, **x2** and $\mathbf{y}$, respectively. Given, the QM output of **x1** can be improved to have lower residuals, $SSR_{\mathbf{x1}} = \sum_i \varepsilon_{\mathbf{x1}}^2(i)/n > 0$. Consequently, $\exists j$, such that

$R_{\mathbf{x1}}(j) \neq O(j)$. Let $\mathbf{x2}(i) = \mathbf{x1}(i)\ \forall i$, where $R_{\mathbf{x1}}(i) = O(i)$ and $\exists k$, such that $R_{\mathbf{x1}}(k) \neq O(k)$ and $R_{\mathbf{x2}}(k) = O(k)$ or $R_{\mathbf{x2}}(k) = R_{\mathbf{x1}}(k)$. Therefore, $\forall i, \varepsilon^2_{\mathbf{x1}}(i) \geq \varepsilon^2_{\mathbf{x2}}(i)$. And since $SSR = \sum \varepsilon_i^2$, we have $SSR_{\mathbf{x2}} \leq SSR_{\mathbf{x1}}$. Thus proving that it SSR of QM output can be improved by correcting the ranks of those data points that do not have the same rank as its corresponding response data point. ♢

Improving the rank correlation of predictor variable to perfectly match the response variable would result in QM output having zeros SSR.

Proposition 4. *The residual error obtained from QM is zero when there is perfect rank correlation between predictor and response variable.*

Proof. Let the elements of R and O be the quantile positions of the data points in $\mathbf{x}$ and $\mathbf{y}$ respectively. Let $\varepsilon_i = |F^{-1}_{\mathbf{y}}(O_i) - F^{-1}_{\mathbf{y}}(R_i)|$ be the residual error of ith data point. Therefore, $SSR = \varepsilon_i^2/n$. Given a perfect rank correlation ($\Gamma = 1$) between predictor and response variable, we have $\forall i$, $(R_i = O_i)$. Consequently, $\varepsilon_i = |F^{-1}_{\mathbf{y}}(O_i) - F^{-1}_{\mathbf{y}}(O_i)| = 0$. Therefore, $SSR = \sum_i \varepsilon_i^2/n = 0$. ♢

Hence, we propose a framework that improves on the ordering of the predictor variables to better match the response variable in order to minimize the SSR of a QM output.

4 Multi-Output Contour Regression Framework(MCR)

Since QM and regression-based approaches have their own distinct advantages which have been successfully exploited in a hybrid manner by approaches such as CR, we propose a framework that extends the intuition behind hybrid approaches that exploits the unique advantages of both QM and regression, to work in a multi-output setting. The approach uses a position regularized regression function $h(\mathbf{x}, \hat{\Omega})$ that prioritizes matching the positions of output to best match the positions of the observed response data. This step is followed by correcting the geometric quantiles of the output from the previous step to match the observed response data using the intuition of QM. This hybrid approach addresses the limitation of QM regarding the number of predictor variables that may be used as well as requirement of the predictor variables being highly correlated to the response variable. We further enhanced the hybrid approach to be flexible enough to work in a multi-output setting so as to be able to capture the multi-output associations that are often ignored.

To prioritize improving the positions of the output, the proposed multi-output contour regression (MCR) framework learns the regression function $h(\mathbf{x}, \hat{\Omega})$. The regression function $h(\mathbf{x}, \hat{\Omega})$ consists of two components. The first component is similar to conventional regression loss function where the data matrix is made to regress with respect to the observed response variable. This component emphasizes minimizing residual error of the regression function.

The second component of $h(\mathbf{x}, \hat{\Omega})$ is the position regularizer that helps improve rank correlation of $h(\mathbf{x}, \Omega)$ and $\mathbf{y}$. At a first glance, one would expect the second term to be regressing on the position of the data points. Instead of regressing

on the position of the data points, we regress on the geometric quantiles of the data points obtained by inverse mapping their positions to the output response space. This is done so that the position regularizer assigns a larger penalty to those data points whose position when incorrectly estimated, results in a larger minimum residual errors. To accomplish this, the data matrix is made to regress on $\mathbf{z}_{\hat{Y}Y}$, where,

$$\hat{\mathbf{z}}_{\hat{Y}Y}(y) = \mathbf{p}_{\hat{Y}}^{-k}(\mathbf{p}_{Y}^{k}(\mathbf{y})) \tag{3}$$

is the geometric quantile value in the $h(\mathbf{x}, \hat{\Omega})$ regression output space that corresponds to the position of the observed response variable y.

The regression function of MCR is shown in Equation (4),

$$\min_{\Omega} \sum_{i=1}^{n} (\gamma \mathcal{L}(h(\mathbf{x}_i, \Omega), \mathbf{y}_i) + (1-\gamma)\mathcal{L}(h(\mathbf{x}_i, \Omega), \mathbf{z}_{\hat{Y}Y})) \tag{4}$$

where $0 \leq \gamma \leq 1$ is a user defined parameter that may be used for either prioritizing fidelity of regression accuracy or its position correlation.

$\mathcal{L}$ can be any generic loss function such as ordinary least square (that multiple linear regression adopts), or quantile mapping (if certain quantiles are to be prioritized overs others, such as in the case of a heavy tail distribution).For instance, when the loss function $\mathcal{L}$ is ordinary least square, Equation 4 takes the form

$$\min_{\Omega} \sum_{j=1}^{q} \sum_{i=1}^{n} (\gamma(\mathbf{x}_i^T \Omega_j - \mathbf{y}_i)^2 + (1-\gamma)(\mathbf{x}_i^T \Omega_j - \mathbf{z}_{\hat{Y}Y})^2)$$

which corresponds to the following matrix form

$$\hat{\Omega} = \arg\min_{\Omega} \; tr(\gamma(\mathbf{X}\Omega - \mathbf{Y})^T(\mathbf{X}\Omega - \mathbf{Y}) + (\mathbf{X}\Omega - \mathbf{Z}_{\hat{Y}Y})^T(\mathbf{X}\Omega - \mathbf{Z}_{\hat{Y}Y}))$$

The regression parameters $\hat{\Omega}$ is learnt in an iterative manner. At each iteration, the regression output space from the previous iteration is used to compute $\mathbf{z}_{\hat{Y}Y}$ in the second component of the regression function $h(\mathbf{x}, \hat{\Omega})$. For the very first iteration, the regression output space is that of regular multiple linear regression.

Once $h(\mathbf{x}, \hat{\Omega})$ is learnt, the MCR prediction for a given data point $\mathbf{x}$ having corresponding observed multi-output response $\mathbf{y}$ and a regression estimation of $\hat{\mathbf{y}} = h(\mathbf{x}, \hat{\Omega})$ is obtained by inverse geometrically quantile mapping $\mathbf{p}_{\hat{Y}}^{k}(\hat{\mathbf{y}})$ to its corresponding value in the observed response variable space, to give the MCR prediction $\hat{\mathbf{z}}_{Y\hat{Y}}$,

$$MCR : \hat{\mathbf{z}}_{Y\hat{Y}} = \mathbf{p}_{Y}^{-k}(\mathbf{p}_{\hat{Y}}^{k}(h(\mathbf{x}, \hat{\Omega}))) \tag{5}$$

where, $\mathbf{p}_{Y}^{-k}(\mathbf{p}_{\hat{Y}}^{k}(\hat{y}))$ maps the stationary geometric quantile position of $h(\mathbf{x}, \hat{\Omega})$ to its corresponding data point in $\mathbf{Y}$.

To summarize, multi-output contour regression (MCR) performs multi-output regression of the predictor variables such that the position of its output is highly correlated with respect to position of the observed response variable, thereby reducing position errors of the multi-output regression results. This multivariate regression output is then mapped to its corresponding geometric quantile counterpart in the observed multi-output response space using geometric quantiles. The rationale behind using the regularized regression results, prior to performing multi-output geometric quantile mapping in MCR, is to improve on SSR by increasing the correlation among the multivariate ranks of the predictors and response variable.

4.1 Estimating Inverse Geometric Quantile Position

The value $\hat{\mathbf{z}}(\mathbf{p})$ that corresponds to a given geometric quantile position $\mathbf{p}$, in a multivariate distribution F_Y i.e., $\mathbf{p}_Y(\mathbf{p})$, is empirically computed by minimizing the generalized multivariate quantile loss function [9]

$$\hat{\mathbf{z}}(\mathbf{p}) = \arg\min_{\mathbf{z}\in\Re^q} \sum_{i=1}^{n} (\|\mathbf{y}_i - \mathbf{z}| + < \mathbf{p}, \mathbf{y}_i - \mathbf{z} >) \tag{6}$$

where, $\mathbf{p} \in \Re^q$ and $< ., . >$ denotes the Euclidean inner product. So long all the values of y_i does not fall on the same line, $\hat{\mathbf{z}}(\mathbf{p})$ will be unique for a given $\mathbf{p}$ for $q \geq 2$ [9]. Algorithms such as Newton-Raphson's method can be used to solve the above loss function geometric quantile $\hat{\mathbf{z}}(\mathbf{p})$ using the following update $\hat{\mathbf{z}} \leftarrow \hat{\mathbf{z}} - \frac{\delta}{\delta'}$ where, $\delta = \sum_{i=1}^{n}((n\kappa)\mathbf{p} - \|\mathbf{z} - \mathbf{y}_i\|^{-1}(\mathbf{z} - \mathbf{y}_i))$

$$\delta' = \sum_{i=1}^{n} \|\mathbf{z} - \mathbf{y}_i\|^{-1}(I_q - \|\mathbf{z} - \mathbf{y}_i\|^{-2} \times (\mathbf{z} - \mathbf{y}_i)(\mathbf{z} - \mathbf{y}_i)^T)$$

For a univariate distribution, F_Y, it can be easily shown that equation (6) boils down to the same loss function used to identify the αth regression quantile in a linear regression setup for quantile regression [15], where $0 < \alpha < 1$ and $p = 2\alpha - 1$. i.e, $\sum_1^n(|y_i - z| + p(y_i - z))$ is minimized for z that corresponds to the αth quantile of Y.

4.2 Alternate Approximation-Based Approach for MCR

If one can make the assumption that given the position ($\mathbf{p}$) of a test data point ($\mathbf{y}^{test}$) that belongs to the distribution F_Y, and $\exists \mathbf{y}_i \in \mathbf{Y}$ such that $\mathbf{y}^{test} \simeq \mathbf{y}_i$, then the search space for $\hat{\mathbf{z}} = \mathbf{y}^{test}$ can be limited to data points in $\mathbf{Y}$.

Given that the search space for $\hat{\mathbf{z}}$ is finite it will not always possible to find the exact same point in F_Y using the loss function δ, as it returns a vector. Alternatively, the following range bound approximation that is equivalent to Equation 6, can be used to find the best solution [11,9].

$$\arg\min_{\mathbf{z}} \sum_{i=1}^{n} \{\| \mathbf{y}_i - \mathbf{z} \| + \frac{1}{\kappa}(\mathbf{y}_i - \mathbf{z})^T\mathbf{p}\} \tag{7}$$

where κ in the scaling factor chosen in Equation (4).

As shown in the experiment section, there was only a marginal performance deterioration in the solution obtained from the above approximation, due to sufficient amount of training data points. Another approximation approach with even less tighter bounds than Equation 7, having $O(n)$ time complexity is to use the following Euclidean approximation.

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{y}_i}((\mathbf{p} - \mathbf{p}_Y(\mathbf{y}_i))(\mathbf{p} - \mathbf{p}_Y(\mathbf{y}_i))^T) \tag{8}$$

The R-limited approximation approach (Equations 7) as well as the Euclidean approximation approach (8) show considerable improvement in the computation time across varying training size (Figure 2.a) and test size (Figure 2.b), with minimum deterioration in terms of accuracy of the inverse geometric quantile positions estimated.

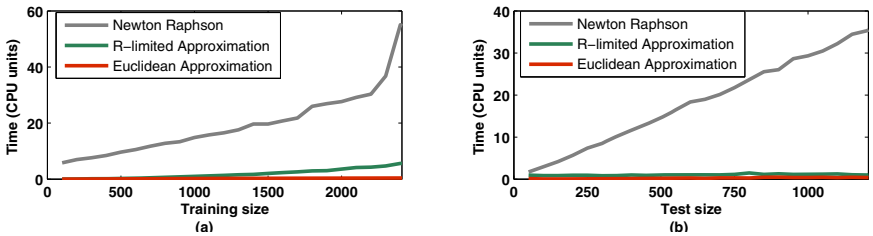

Fig. 2. Relative computation time of the various approximation-based approaches for estimating inverse geometric quantile positions

5 Experimental Results

The objective of the experiments was to evaluate the ability of MCR in replicating the associations among multiple climate response variables while minimizing sum square residuals.

All the algorithms were run using climate data obtained at fourteen weather stations in Michigan, USA. The response variables used were maximum temperature, minimum temperature, and the total precipitation for each day spanning twenty years. The predictor variables used in this study are simulated climate data obtained from Regional Climate models (RCM) that best correspond to the observed response variables at each of the fourteen weather stations. Three different RCM data sets for each of the climate stations were obtained from North American Regional Climate Change Assessment Program (NARCCAP) [1]. The three RCMs used are the Canadian Regional Climate Model (CRCM), the Weather Research and Forecasting Model (WRFG) and the Regional Climate Model Version-3 (RCM3). For the purpose of the experiments, there were a total of 126 data sets with univariate response variables, 126 data sets with bivariate responses and 42 data sets with trivariate responses.

5.1 Experimental Setup

Twenty year of predictor and response data, spanning the years 1980-1999 was split into two parts for training and testing. For the purpose of the evaluation of the relative skill in preserving associations among the multi-output responses, popular regression and quantile mapping approaches such as MLR, Ridge regression (Ridge), QM, EDCDFm, MOR, CR, BQM as well as ad-hoc approaches that sequently combine regression and quantile mapping approaches were used as baseline. An example of the ad-hoc baseline approach used is MOR in combination with BQM (RBQM) and MLR and QM (RQM). γ was set to 0.5 for all experiments. For CR and MCR based experiments, the maximum number of iterations was set to ten.

After discarding the missing values, each experiments run for each of the stations, across all the data sets, had a minimum of one thousand training and test data points. All the results provided in the following section are on test data (out-of-sample results). Kendall τ rank correlation and Spearman ρ rank correlation were the two rank correlation metrics used for evaluation univariate rank correlation. In the following experiment section, we included results of only one of the two rank correlation metrics, when their results were very similar. Root mean square error (RMSE), was used as a metric to compare the performance of the various approaches evaluated in terms of its output residual errors. Two dimensional and three dimensional scatter plots were used to visualize the relative skill of the various approaches in preserving the associations among the multi-output responses.

5.2 Results

Univariate MCR. For academic reasons, the rank correlation of the various response variables were computed in a single output MCR setting using Kendall τ rank correlation and Spearman ρ rank correlation. It was found that across all the different data sets and stations and response variables (i.e, 126 datasets), MCR consistently improved the rank correlation across both rank correlation metrics. For the purpose of comparing the intra-performance of datasets that shared similar response variables, the 126 individual data sets that corresponded to univariate response data were grouped into nine larger data sets.

Figure 3 is a box plot representing the percentage of stations in each of the nine data sets where the rank correlation regularizer used in Equation 4, improved rank correlation and reduced residuals when compared to baselines approaches.

The box plot in Figure 4 shows that in spite of MCR's reported improvement across majority of stations in terms of τ and RMSE, for both regression and quantile mapping based approaches, the improvement was not significant when compared to the regression based approaches. However, the rank correlation regularizer showed a significant improvement in terms of RMSE at each station when compared to the corresponding quantile mapping based approaches.

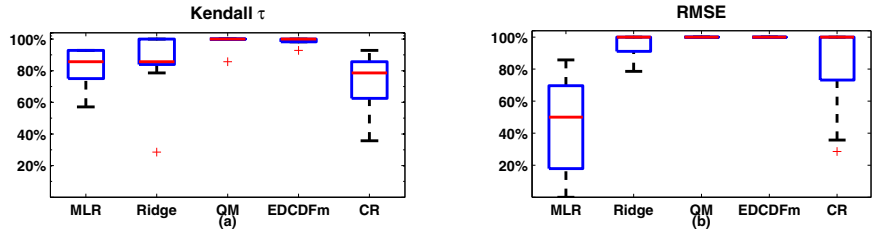

Fig. 3. Box plot of the percentage stations where MCR showed improvement over single output baselines, in terms of Kendall τ rank correlation and RMSE, across all RCM's and variables

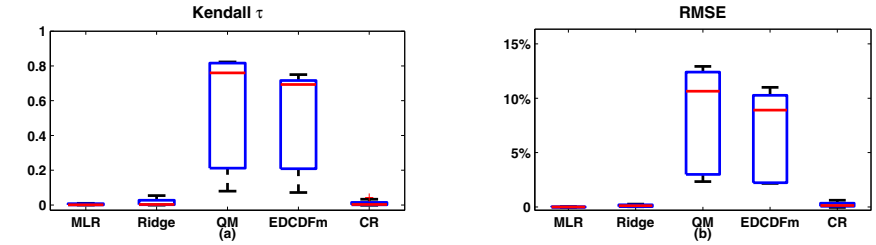

Fig. 4. Box plot of MCR's improvement over baseline approaches in terms of Kendall τ rank correlation and RMSE, across all RCM's and variables

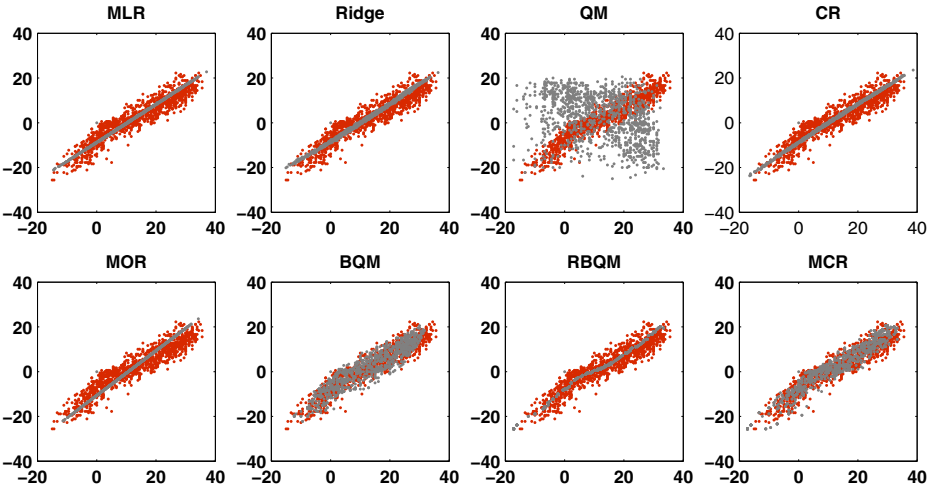

Fig. 5. Scatter plot portraying the fidelity of forecast values of various approaches replicating the observed associations among the bivariate temperature response variables

Bivariate MCR. Bivariate modeling for all the combinations of bivariate response variables were evaluated. As shown in Figure 5, MCR performed best in replicating both the bivariate associations and minimizing SSR, although BQM performed as well in terms of replicating the bivariate associations. Regression based approaches (both SOR and OMR) fared poorly in preserving associations in the 2D space, while single output quantile mapping based approaches, were very sensitive to correlation of the predictor variables with response resulting in poor bivariate associations in spite of replicating the marginal distributions of the individual responses very well.

Table 3. Performance of bivariate MCR over baseline approaches

Data set	RMSE						Kendall τ					
	% of stations outperformed baseline			Avg.improvement across stations over baseline			% of stations outperformed baseline			Avg.improvement across stations over baseline		
	MOR	QM	BQM	MOR	QM	BQM	MOR	QM	BQM	MOR	QM	BQM
$WRFG_1$	29	100	100	-0.06	0.18	0.17	64	100	100	0.03	0.40	0.41
$WRFG_2$	07	100	100	-0.08	0.16	0.16	79	100	100	0.04	0.38	0.39
$WRFG_3$	00	100	100	-0.07	0.31	0.30	0	100	100	-0.01	0.75	0.67
$CRCM_1$	93	100	100	0.06	0.25	0.25	100	100	100	0.13	0.52	0.53
$CRCM_2$	71	100	100	0.03	0.23	0.23	100	100	100	0.12	0.49	0.52
$CRCM_3$	07	100	100	-0.02	0.35	0.34	14	100	100	-0.01	0.78	0.73
$RCM3_1$	43	100	100	-0.02	0.20	0.20	79	100	100	0.06	0.46	0.46
$RCM3_2$	36	100	100	-0.03	0.19	0.18	79	100	100	0.06	0.47	0.45
$RCM3_3$	00	100	100	-0.07	0.31	0.30	0	100	100	-0.01	0.81	0.78

In terms of residuals, MCR had considerably lower residuals when compared of the various quantile mapping baseline approaches as shown in Table 3. But as expected, MCR showed marginal increase in residuals when compared to the respective SOR and MOR based approaches.

Trivariate MCR. The performance of modeling the association among three response variables was also evaluated and is shown in Figure 6. The performance is compared against single output, and multiple output models. We also use as a baseline, an trivariate extension of the bivariate BQM approach, as an additional baseline. Along with MCR, the trivariate extension of BQR fared best in replicating the observed associations among three variables when compared to the baseline approaches.

Additionally, MCR was also able to improve upon its BQM counterpart in terms of reduction of residuals. MCR produced lower RMSE for all the station across all the tri-variate datasets with an average reduction of RMSE in excess of 10%. The average improvement of the three variables in terms of rank correlation τ was found to be 0.41.

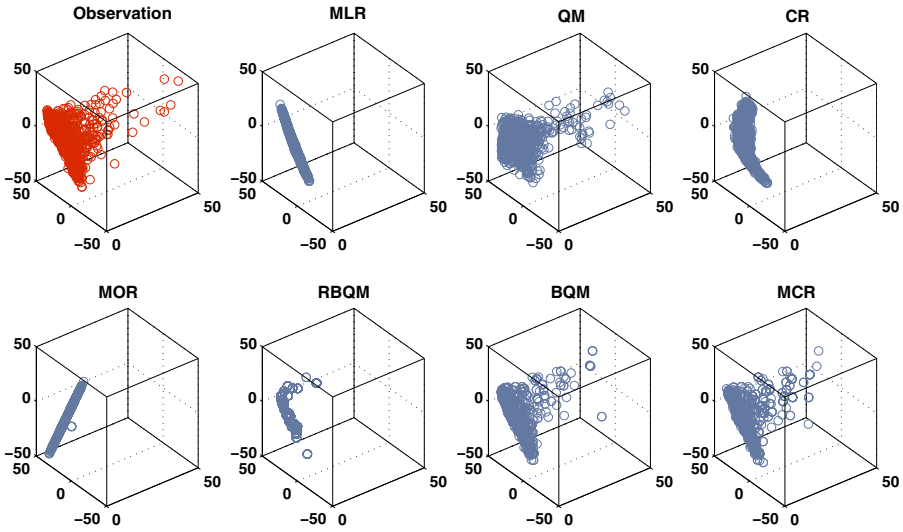

Fig. 6. Three dimensional scatter plot of the observed associations among maximum temperature, minimum temperature and precipitation as well as the respective forecasts made by the various single output and multiple output approaches

6 Conclusions

We present a multi-output regression framework that preserves the general association patterns among multiple response variables while minimizing the overall residual errors by coupling regression and geometric quantile mapping. The paper demonstrates the effectiveness of the framework in significantly reducing residuals while preserving the joint distribution of the multi-output variables, over the baseline approaches in all the climate stations evaluated.

Acknowledgments. This work is supported by NSF Award CNH 0909378. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. North American Regional Climate Change Assessment Program, `http://www.narccap.ucar.edu/`
2. Abraham, Z., Tan, P.-N.: Distribution regularized regression framework for climate modeling. In: SIAM International Conference on Data Mining, SDM 2013 (2013)
3. Alvarez, M., Lawrence, N.D.: Sparse convolved gaussian processes for multi-output regression. In: Advances in Neural Information Processing Systems 21, pp. 57–64 (2009)

4. Balasubramanian, K., Lebanon, G.: The landmark selection method for multiple output prediction. In: International Conference on Machine Learning, ICML (2012)
5. Blaschko, M.B., Lampert, C.H.: Learning to localize objects with structured output regression. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 2–15. Springer, Heidelberg (2008)
6. Bontempi, G.: Long term time series prediction with multi-input multi-output local learning. In: Proceedings of the 2nd European Symposium on Time Series Prediction (TSP), ESTSP 2008, pp. 145–154 (2008)
7. Breiman, L., Friedman, J.H.: Predicting multivariate responses in multiple linear regression. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 59(1), 3–54 (1997)
8. Buja, A., Logan, B., Reeds, J., Shepp, L.: Inequalities and positive-definite functions arising from a problem in multidimensional scaling. The Annals of Statistics, 406–438 (1994)
9. Chaudhuri, P.: On a geometric notion of quantiles for multivariate data. Journal of the American Statistical Association 91(434), 862–872 (1996)
10. Hastie, T., Tibshirani, R., Friedman, J.J.H.: The elements of statistical learning, vol. 1. Springer, New York (2001)
11. He, X., Yang, Y., Zhang, J.: Bivariate downscaling with asynchronous measurements. Journal of Agricultural, Biological, and Environmental Statistics 17(3), 476–489 (2012)
12. Izenman, A.J.: Reduced-rank regression for the multivariate linear model. Journal of Multivariate Analysis 5(2), 248–264 (1975)
13. Jakob Themeßl, M., Gobiet, A., Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. International Journal of Climatology 31(10), 1530–1544 (2011)
14. Kim, S., Xing, E.P.: Tree-guided group lasso for multi-task regression with structured sparsity (2010)
15. Koenker, R.: Quantile Regresssion. Wiley Online Library (2005)
16. Marden, J.I.: Positions and qq plots. Statistical Science 19(4), 606–614 (2004)
17. O'Brien, T., Sornette, D., McPherron, R.: Statistical asynchronous regression: Determining the relationship between two quantities that are not measured simultaneously. Journal of Geophysical Research 106(A7), 13247–13259 (2001)
18. Pérez-Enciso, M., Tenenhaus, M.: Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (pls-da) approach. Human Genetics 112(5-6), 581–592 (2003)
19. Piani, C., Weedon, G., Best, M., Gomes, S., Viterbo, P., Hagemann, S., Haerter, J.: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. Journal of Hydrology 395(3), 199–215 (2010)
20. Vazquez, E., Walter, E.: Multi-output support vector regression. In: 13th IFAC Symposium on System Identification, pp. 1820–1825 (2003)