In recently, there are some results with intelligent, computational intelligence, machine learning and soft computing methods for weather forecasting problem. However, these methods to deal with weather forecast problem that only consider single-valued weather data. Hence, in this sub-proposal we considered histogram-valued/distribution-valued weather data that provided different viewpoint in weather forecast problem. In the second year, we extended to uses the distributional variable to replace histogram-valued variable. We proposed an improved norm-distribution principal component analysis (INDPCA). Besides, we also derived some different distances measure; namely, Wasserstein distance and generalized Wasserstein distance on the distributional variable data analysis.

Methods and Results

For the dataset of this paper can be represented as $X_{m \times p}$, each element $X_{q}$ must be a single value. The Distribution Valued of $X_{y}$ to $X_{p}$ is represented as: $N_{p} = (N_{1},...,N_{3},...,N_{p})$. With $F_{(t)}$ its cumulative distribution function (cdf) and with $G_{(t)}$ (for $t \in [0, 1]$) the corresponding qf (i.e., the inverse of the cdf) [3]. The $W_{n}$ Wasserstein distance between two distribution functions can be expressed as:

$$d_{W_{n}}(N_{i},N_{j}) = \left[ \int_{0}^{1} |F_{i}^{-1}(t) - G_{j}^{-1}(t)|^{n} \, dt \right]^\frac{1}{n}$$

Ipino and Romano (2008) proved also a general formulation of the Wasserstein distance [2]; if $F_{i}$ and $F_{j}$ are the distribution functions of two random variables $f$ and $g$, we consider three component: Location, Size and Shape, Wasserstein distance can be written as:

$$d_{Wass}^{2} = (\mu_{f} - \mu_{g})^{2} + (\sigma_{f} - \sigma_{g})^{2} + 2\sigma_{f}\sigma_{g}(1 - \rho_{QQ}(F,G))$$

where

$$\rho_{QQ}(F,G) = \frac{\int_{0}^{1} F^{-1}(t) - G^{-1}(t) \, dt}{\sqrt{\int_{0}^{1} F^{-1}(t) - \mu_{f} \, dt \cdot \int_{0}^{1} G^{-1}(t) - \mu_{g} \, dt}}$$

is the correlation of the quantiles of the two distributions as represented in a classical QQ plot.

Fig. 1. The weather distribution-valued data.

Table 1. Wasserstein distance between cities of Taiwan on temperature distribution-valued data in 2016.

<table>
<thead>
<tr>
<th>City</th>
<th>Taipei</th>
<th>Hualien</th>
<th>Chiayi</th>
<th>Kaohsiung</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taipei</td>
<td>0</td>
<td>2.8544</td>
<td>44.4826</td>
<td>31.8705</td>
</tr>
<tr>
<td>Hualien</td>
<td>0</td>
<td>6</td>
<td>42.6143</td>
<td>29.8602</td>
</tr>
<tr>
<td>Chiayi</td>
<td>0</td>
<td>6</td>
<td>42.6143</td>
<td>15.5359</td>
</tr>
<tr>
<td>Kaohsiung</td>
<td>0</td>
<td>6</td>
<td>42.6143</td>
<td>15.5359</td>
</tr>
</tbody>
</table>

Conclusion

This year, we extended to uses the distributional variable to replace histogram-valued variable. We proposed an INDPCA. Besides, we also derived some different distances measure; namely, Wasserstein distance and generalized Wasserstein distance on the distributional variable data analysis.

References