Statistical downscaling of precipitation using inclusive multiple modelling (IMM) at two levels

Sina Sadeghfam, Rahman Khatibi, Tara Moradian and Rasoul Daneshfar
Department of Civil Engineering, Faculty of Engineering, University of Maragheh, Maragheh, East Azerbaijan, Iran

Abstract
Topical research on hydrological impacts of climate change in terms of downscaling of monthly precipitation is investigated in this paper by formulating an inclusive multiple modelling (IMM) strategy. IMM strategies manage multiple models at two levels and the paper uses statistical downscaling model, Sugeno fuzzy logic and support vector machine at Level 1 and feeds their outputs to a neuro-fuzzy model at Level 2. In the downscaling stage, large-scale NCEP (National Centres for Environmental Prediction)/NCAR (National Centre for Atmospheric Research) data for a station with local data record from 1961 to 2005 are used for training and testing Level 1 models, which are found to be ‘fit-for-purpose’, but the variations between them signify some room for improvements. The model at Level 2 combines outputs of those at Level 1 and produces Level 2 results, which are over the Level 1 models in terms of dispersion of residual errors. In this way, IMM provides a more defensible modelling strategy for application in the projection stage. The comparison between observed and projected precipitation indicates that precipitation will be likely to reduce compared with observed precipitation in cold seasons (October–February), but the projected precipitation will be likely to increase slightly in wet seasons (April and May).

Key words: artificial intelligence models, downscaling, inclusive multiple modelling (IMM), precipitation, projection

Highlights
- Inclusive multiple modelling (IMM) is formulated for the downscaling of precipitation.
- IMM manages multiple regression and artificial intelligence models at two levels.
- Performances of IMM are improved over Level 1 models in the downscaling stage.
- IMM provides a more defensible strategy for application in the projection stage.

1. Introduction
Enhancing the accuracy of the models on statistical downscaling of precipitation is investigated in this paper using inclusive multiple modelling (IMM) practices, where statistical downscaling refers to transforming large-scale predictor variables to local climate variables referred to as predictand(s), e.g., precipitation by a statistical tool. There is no single way of modelling a predictand variable, as the possibilities are large in a pluralistic modelling environment and the paper tests the application of IMM practices. Khatibi et al. (2020) and Khatibi & Nadiri (2020) present these practices as a way of maximising the extracted information or enhance accuracies of local models through a selection of a limited number of models without exhaustive testing of all of the available models. IMM is outlined further below and is implemented through formulating strategies to enhance the correlation between large-scale predictors with local precipitation (predictand) at two levels: at Level 1, a number of generally available models are constructed using predictors and the predictand(s) and at Level 2, yet another model is constructed, which reuses the predictions of the models at Level 1 as its inputs.

General circulation models (GCMs) provide simulations of large-scale weather variables in the future for climate studies. Climate change impacts water resources and is investigated in terms of climate variables on a local scale by downscaling climate variables from large-scale to local-scale variables. The following downscaling approaches are widely used: (i) dynamical or physical downscaling and (ii) statistical or empirical downscaling. Dynamical downscaling extracts large-scale information...
from GCMs at low resolutions and maps them onto higher resolution models known as the limited area method (LAM) or the regional climate model (RCM) (e.g., Ishida et al. 2020). Extracting local information using LAM or RCMs requires complex design and computational time, which limits their use (Anandhi et al. 2008). Statistical downscaling seeking an empirical relationship between climate predictors at large-scale climate variables and the predictand of local climate variables in the historical period. Its main drawback is its assumption, in which the derived relationship between predictors and predictand in the historical period is extrapolated to future periods, although they are simple and used widely.

Statistical downscaling models (SDSMs) are categorised into (i) statistical or regression models and (ii) artificial intelligence (AI)-based models. Statistical models include SDSMs (Wilby et al. 2002) and LAS-WG (Semenov & Barrow 1997), which use multiple linear regression (MLR) and are widely employed to downscale daily and monthly precipitation and temperature (e.g., Chen et al. 2010; Pervez & Henebry 2014; Gulacha & Mulungu 2017). Overall, they seek a complex relationship between predictors and the predictand for precipitation downscaling. Notably, SDSM performances in monthly scales are reportedly better than their daily scale as extreme events lie beyond the range of regression models (Pervez & Henebry 2014), and hence a further focus is given below on both statistical and AI models.

Previous studies in statistical downscaling have addressed various aspects of the problem, which include (i) using statistical and AI models to identify the relation between predictors and the predictand(s) for a local area (e.g., Chen et al. 2010; Li et al. 2020); (ii) using different screening techniques to select predominant predictors among a large number of predictors (e.g., Nourani et al. 2019); (iii) evaluating different GCMs or RCMs for statistical downscaling (e.g., Sing et al. 2019; Li et al. 2020) and (iv) combining the results of different models (e.g., Su et al. 2019). Table 1 presents an outline in terms of their variations, their downscaling models and predictor screening techniques.

Research activities in the first group include Chen et al. (2010), who use SDSM, multiple regression and support vector machine (SVM) for daily precipitation downscaling. They used discriminant analysis and support vector classification for classifying days as wet and dry and utilised multiple regression and SVM to estimate precipitation in wet days. Their results show that SVM produces more reasonable results, although SDSM has a better performance for cases when daily precipitations are small. Hashemi et al. (2011) compared SDSM with gene expression programming (GEP) for downscaling daily precipitation; they observe that GEP offers a slight improvement and some advantages including automatic screening of the predictors. Employing artificial neural network (ANN), GEP, SVM and relevance vector machine (RVM) to downscale monthly precipitation, Sachindra et al. (2018) observe that AI models have better performances in estimating mean precipitation but underestimate the maximum and standard deviation of precipitation series. They observe that SVM and RVM score the highest performance metrics but recommend RVM or ANN over SVM or GEP for flood prediction studies and RVM for drought studies.

The activities in the second group included screening techniques to select predominant predictors. Hammami et al. (2012) developed a penalised regression method, referred to as the least absolute shrinkage and selection operator (LASSO) to identify the dominant predictors, and the results indicated that LASSO performed better than the classic approach. Using ANN to downscale monthly precipitation, Nourani et al. (2019) investigate three screening techniques to select dominant large-scale predictors, which include decision tree, mutual information and correlation coefficient (CC). They report that training ANN performs better when predictors are selected by incorporating decision tree. Teegavarapu & Goly (2018) investigate stepwise regression model (SRM), mixed-integer nonlinear programming (MINLP) and ANN to select dominant predictors and report that MINLP and ANN improve downscaling performances but selecting predominant predictors is site-specific. Jafarzadeh et al. (2021) identify dominant predictors by four different algorithms and observe that among them the Bayesian theorem and stepwise regression performed better than others.

The third group of activities evaluates or ranks different GCMs or RCMs. Ahmed et al. (2020) report on downscaling monthly precipitation, maximum and minimum temperatures using ANN, K-nearest neighbour (KNN), SVM and RVM. They rank 15 GCMs and report HadGEM2-AO as the most skilled model and conclude that KNN and RVM exhibit higher performance than SVM and ANN. Employing linear regression for downscaling precipitation obtained by predictors belonging to various GCMs and RCMs, Sing et al. (2019) report that their results reveal the underestimation and overestimation of GSMS and RCMs, respectively, but show RCM to perform closer to observed precipitation.

The fourth group of activities addresses topical research on combining different results of statistical downscaling. Using Bayesian model averaging (BMA) to combine different regression models, Su et al. (2019) incorporate different predictors and compare the results with SRM. They conclude that BMA performs better than SRM since using too many predictors in a regression model decreases the predictive power, and hence BMA overcomes this limitation. Downscaling daily
Table 1 | Over review of previous studies

| Researcher(s) | Study area | Climate model | Statistical downscaling | Predictor screening technique | Time period | Time scale | Downscaled parameter | Transformation |
|---------------|------------|---------------|--------------------------|------------------------------|-------------|------------|--------------------|---------------|
| Chen et al. (2010) | Shih-Men Reservoir basin (Taiwan) | Had-CM3<sup>a</sup> | SDSM<sup>b</sup> | Kolmogorov-Smirnov test | Daily | 1964–1990 | ✓ | C | Norm. – | A2 B2 |
| Hashemi et al. (2011) | Chateau Watershed (New Zealand) | Had-CM3 | GEP<sup>c</sup> | – | Daily | 1961–1990 | ✓ | – | – – | – – |
| Pervez & Henebry (2014) | Two river basins (South Asia) | CGCM3<sup>d</sup> | CC | Daily | 1988–1997 | ✓ | Fourth Root | – A2 | A1B B2 |
| Gulach & Mulungu (2017) | The Wami-Ruvu River Basin (Tanzania) | Had-CM3 | Partial correlation | Monthly | 1961–1975 | ✓ | MacT MinT UC | – – | A2 B2 |
| Nourani et al. (2019) | Tabriz station (Iran) | Can-ESM2<sup>e</sup> | Decision tree Mutual information | CC | Monthly | 1950–1991 | ✓ | MeanT | Std – | RCP4.5 RCP8.5 |
| Sachindra et al. (2018) | Victoria State (Australian) | NOAA<sup>f</sup> | GEP | – | – | 1971–2012 | ✓ | – | Std | – – |
| Su et al. (2019) | Heshe River basin (China) | NOAA | BMA<sup>g</sup> | PCA | Monthly | 1951–2000 | ✓ | – | – | – | RCP8.5 RCP4.5 |
| Sang et al. (2019) | India | GCMs including Can-ESM2 RCMs including CNRM-CM5<sup>h</sup> | Linear regression | – | – | 1961–1992 | ✓ | MacT MinT | – – | – – |
| Ahmed et al. (2020) | Pakistan | 15 GCMs including HadGEM2 (as the best) | ANN | Monthly | 1961–1992 | ✓ | MacT MinT | – | – | – |
| Li et al. (2020) | Ontario (Canada) | GCM including Can-ESM2, EC̶EARTH RCM including CanRCM4, CRCM5<sup>i</sup> | Arithmetic mean | Daily | 1980–1986 | ✓ | – | – | – |

<sup>a</sup>Hadley Centre Coupled Model, Version 3.
<sup>b</sup>Statistical downscaling model.
<sup>c</sup>Support vector machine.
<sup>d</sup>Support vector classification.
<sup>e</sup>Support vector regression.
<sup>f</sup>Gene expression programming.
<sup>g</sup>Third generation coupled global climate model.
<sup>h</sup>Canadian Center for Climate Modelling and Analysis Earth System Model.
<sup>i</sup>Beijing Normal University Earth System Model.
<sup>j</sup>Multiple linear regression.
<sup>k</sup>Artificial neural network.
<sup>l</sup>National Oceanic and Atmospheric Administration.
<sup>m</sup>Relevance vector machine.
<sup>n</sup>Bayesian model averaging.
<sup{o}</sup>Stepwise regression model.
<sup>p</sup>Centre National de Recherches Météorologique Model 5.
<sup>q</sup>Earth system model low resolution.
<sup>r</sup>Hadley Centre Global Environment Model version 2.
<sup>s</sup>K-nearest neighbour.
<sup>t</sup>European community.
<sup>u</sup>Canadian regional climate models.
temperature using statistical and AI models, Li et al. (2020) report on statistical models, including arithmetic mean and MLR, and combine their predictors with equal and unequal weights, respectively. AI models include long short-term memory and SVM. Their results show that both statistical and AI models perform well but unable to accurately estimate extremes. This group of models is regarded as IMM practices, but the paper develops an IMM strategy for a deeper understanding of the problem.

One of the main gaps in the models for the statistical downscaling of precipitation is the absence of explicit techniques to extract full information from the local data and hence the paper. The application of IMM practices to these models in the monthly time scale at a station is tested for a case study in a region at the southeast coast of the Caspian Sea, northwest of Iran. Notably, limiting the study area to one station makes it easy to study the IMM capability from an elementary basis without undue complications. This simplification is also cited in the literature (e.g., Chen et al. 2010; Nourani et al. 2019) but for other reasons. IMM can combine the results of statistical and AI models at Level 1 with a combiner AI model at Level 2. The novelty of the paper includes the application of IMM to the statistical downscaling of precipitation and the developed models project precipitation into the future, as detailed next.

2. METHODOLOGY

2.1. Critical view of conventional approaches

The various aspects of the past research on statistical or AI-based models on downscaling are outlined above, based on which the paper seeks a deeper understanding of IMM practices. This is in a background where most of the modelling research activities are largely focussed on ranking with the main goal of selecting the better performing one, often referred to as the ‘superior’ model. Khatibi et al. (2020) and Khatibi & Nadiri (2020) refer to this type of modelling practices as exclusionary multiple models (EMMs), and hence these models are better described as ‘fit-for-purpose’ but not superior. One reason is that EMM practices do not take a full benefit of the multiple models but construct them at the expense of the culture of ranking, which transforms models into being ‘the end’. In reality, models ought to be regarded as the ‘means to an end’ of serving as a learning tool stipulated by IMM practices.

The paper takes on board IMM practices, which is driven by formulating modelling strategies to seek ‘defensibility’ of the results by ensuring that the information extracted from site-specific data is enhanced to demonstrable levels. Nadiri et al. (2019) explain that there is already an established mathematical basis for the mean of multiple models to have root-mean-square errors (RMSE less than the average of RMSE of the individual models. Thus, simple averaging of multiple models at Level 1 can be regarded as a set of results at Level 2, suitable for benchmarking the results of a more sophisticated ‘combiner model’ at Level 2. Moreover, the past published results by the authors show that the quality of the results at Level 2 is also improved in various ways and these collectively add to the ‘defensibility’ of the results, and hence, they are ‘fit-for-purpose’ at Level 1 for not using the full potential of the constructed multiple models.

The authors build up evidence for the scope of IMM strategies by testing them in different fields of water and environmental modelling. Past applications include both explicit evidence on IMM (e.g., Karimi et al. 2020; Khatibi & Nadiri 2020; Khatibi et al. 2020; Nadiri et al. 2020) and tacit evidence on IMM (e.g., Khatibi et al. 2017; Ghorbani et al. 2018; Nadiri et al. 2018; Sadeghfam et al. 2019; Moazamnia et al. 2020). These also reflect on the IMM evolution, which goes back to the 1960s, as reviewed by Clemen (1989), although an exhaustive review is beyond the scope of the paper.

An IMM modelling strategy is formulated in the paper, which aims to benefit from multiple models at Level 1 (referred to as base models), but at Level 2, a combiner model reuses base models. Notably, any statistical or AI-based models can be selected and organised as base models and a combiner model, see the illustration in Figure 1. The figure depicts the activities in levels as follows: at Level 0, the available data are reviewed and decisions are made on the model structure and modelling strategies; at Level 1, base models comprise one statistical and two AI-based models: SDSM, Sugeno fuzzy logic (SFL) and SVM and at Level 2, neuro-fuzzy (NF) is selected as the combiner model. These models are only specified in this section in the sense that only so much details are presented that third parties can reproduce the results reported by the paper. All the required procedures were implemented in the MATLAB platform.

2.2. Level 0: pre-processing – selecting dominant predictors

Downscaling involves a selection of predominant predictors, and the paper selects those by the NCEP (National Centres for Environmental Prediction)/NCAR (National Centre for Atmospheric Research) predictors. Generally, NCEP/NCAR
reanalyse data as the historical and large-scale climate variables with the resolution of 2–4° and are globally gridded data involving observations and numerical weather predictions. Table 2 presents the list of 26 predictor variables in NCEP/NCAR.

The selection of predominant predictors is carried out by incorporating a screening technique. CC is a screening technique, which is appropriate for statistical techniques such as SDSM. There are different techniques available in the literature for AI-based models (see Table 1). The paper employs CC for SDSM, as follows:

Predictand at Level 0 = screen(predictor1, predictor2, ... predictor26) (1)

The paper employs decision tree for constructing AI-based models at Level 1 (see Figure 1).

Predictors at Level 0 = M5(predictor1, predictor2, ... predictor26) (2)

where M5 is a decision-tree complexity pruning technique outlined below.

Table 2 | List of predictor variables from NCEP/NCAR and Can-ESM2

| Predictor variable | Name       | Predictor variable | Name       |
|--------------------|------------|--------------------|------------|
| 1,000 hPa wind speed | p1_f       | 850 hPa wind speed | p8_f       |
| 1,000 hPa zonal wind component | p1_u       | 850 hPa zonal wind component | p8_u       |
| 1,000 hPa meridional wind component | p1_v       | 850 hPa meridional wind component | p8_v       |
| 1,000 hPa relative vorticity of true wind | p1_z       | 850 hPa relative vorticity of true wind | p8_z       |
| 1,000 hPa wind direction | p1th      | 850 hPa geopotential height | p850       |
| 1,000 hPa divergence of true wind | p1zh      | 850 hPa wind direction | p8th       |
| 500 hPa wind speed | p5_f       | 850 hPa divergence of true wind | p8zh       |
| 500 hPa zonal wind component | p5_u       | Specific humidity at 500 hPa | s500       |
| 500 hPa meridional wind component | p5_v       | Specific humidity at 850 hPa | s850       |
| 500 hPa relative vorticity of true wind | p5_z       | Specific near-surface humidity | Shum       |
| 500 hPa geopotential height | p500      | Mean temperature at 2 m | Temp.      |
| 500 hPa wind direction | p5th      | Total precipitation | Precp.     |
| 500 hPa divergence of true wind | p5zh      | Mean sea-level pressure | Mslp       |
Reported studies provide evidence on the performance of decision tree in downscaling by AI models (Nourani et al. 2019). It is a supervised AI technique, which is structured in terms of four levels (from the upper to the lower level): root, branches, nodes and leaves. The M5 model was developed by Quinlan (1992) as a classifier technique to understand the relationship between dependent and independent variables. M5 establishes regression equations to each leaf as a small part of data, whereas the classic regression establishes an equation to the whole data. The required procedure comprises (Pal & Deswal 2009): (i) using a split criterion to create a tree model, in which M5 utilises the reduction of standard deviation of each class and (ii) pruning of the branches and substituting them in a regression equation. The dominant predictors can be selected at upper nodes, which have a high priority to predictand. Also, predictors at lower nodes not appeared in the structure of the tree have lower priorities.

2.3. Level 1: base models

2.3.1. Statistical downscaling model

SDSM, developed by Wilby et al. (2002), constructs MLR models as an empirical relationship expressing a predictand at a higher resolution in terms of a set of large-scale predictors at a lower resolution. The set of predictors are generated by stochastic weather generators and assume that the empirical equation in the past remains true for the future. Stochastic weather generator applies the relationships with the probability of precipitation depending on predictors, expressed as follows:

\[
\text{Predictand at Level } 1 = f(\text{screened predictors by Eq. (1)}) \quad (3)
\]

where \( f \) is a function denoting the algorithm related to MLR.

Since precipitation as the predictand has a highly variable nature on a local scale, it cannot be fully described by large-scale predictors. Therefore, stochastic techniques artificially inflate the variance of downscaled precipitation time series (Wilby et al. 2002), where variance inflation refers to the scale variance of the downscaled predictand to achieve better agreement with observed values.

2.3.2 Support vector machine

SVM, developed by (Vapnik 1998), is used as the predictor at Level 1 to downscale predictands identified at Level 0. It is a statistical learning technique and uses a kernel-based learning approach, in which a linear high-dimensional hypothesis space, referred to as feature space, is mapped onto a low-dimensional space using implicit kernel functions. It has similarities with neural networks in terms of weights and bias (Romero & Toppo 2007), in which input data comprise predominant predictors processed at Level 0 by the M5 decision tree; its target data are local precipitation, which also serve as the predictand (see Figure 2). SVM uses the following regression-type equation:

\[
\text{Predictand at Level } 1 = f'(\text{M5 pruned predictors by Eq. (2)}) \quad (4)
\]

where \( f' \) is a function denoting the algorithm related to SVM. In the study, radial basis function (RBF) was employed as the kernel function, and related parameters for RBF are obtained by the least-square procedure as discussed by Suykens et al. (2002).

2.3.3. Sugeno fuzzy logic

SFL is a widely used AI technique, for more details see Takagi & Sugeno (1985) and for more details on the author’s implementation, see Sadeghfam et al. (2018). To the best knowledge of the authors, SFL is yet to be used as an SDSM, but the paper uses it as a downscaling model at Level 1. Its implementation requires the following procedure: (i) a clustering technique is required to identify groupings within the site-specific data to derive rule base automatically and to identify inherent parameters in membership function (MF), in which MF retains the contribution from the original fuzzy logic. The identification of the grouping in the data is through the subtractive clustering (SC) method given by Chiu (1994). SC is processed by using cluster radius to control the number of clusters and fuzzy rules (Chen and Wang 1999). The optimum cluster radius is identified by systematically varying the cluster radius in the range of 0–1 until the optimum value is identified in terms of RMSE.
The fuzzy if-then rule is expressed in the following equation for precipitation downscaling:

\[
\text{Rule } i: \quad \text{If } \begin{cases} \text{Predictor}_1 \text{ belongs to MF}_1 \\ \text{Predictor}_2 \text{ belongs to MF}_2 \\ \text{Predictor}_3 \text{ belongs to MF}_3 \\ \vdots \\ \text{Predictor}_n \text{ belongs to MF}_n \end{cases}, \quad \text{Then } \text{Predictand} = \sum_{i=1}^{n} m_i \times \text{Predictor}_i \tag{5}
\]

where MF is the membership function; \( n \) is the number or predominant predictors and \( m_i \) is the coefficients. The final output \( \text{Out}_j \) is the weighted average of all outputs (aggregation) as follows:

\[
\text{Out}_j = \frac{\sum_k w_{kj} \text{Out}_{kj}}{\sum_k w_{kj}} \tag{6}
\]

where \( w_{kj} \) is the firing strength for rule \( k \) and output \( j \), obtained using the ‘AND’ (minimise) operator.

2.4. Level 2: combiner model

The IMM strategy uses NF, originally given by Jang (1993), as a combiner model at Level 2, in which the downscaled precipitation by the base models at Level 1 is reused as input data for more improved performances. NF determines fuzzy rules by integrating ANN with the fuzzy inference system, such that the parameters of the membership function in SFL are tuned by a hybrid algorithm using neural networks. The implementation of NF involves five layers, but these are detailed in textbooks (Jang et al. (1997). The authors’ implementation is similar to Moazamnia et al. (2019), which may be referred to for further details.

2.5. Performance metrics

The authors do not encourage the use of superlatives such as superior models in the inter-comparison of models based on anecdotal test cases. It is unlikely to be any superior model, but one model may explain a set of local data better than others. As discussed by Khatibi et al. (2020) and Khatibi & Nadiri (2020), it is more appropriate to express model performances in terms of ‘fit-for-purpose’ for those based on EMM practices and as ‘defensible’ for those based on IMM practices. The ultimate aim is to ensure that any information contained in the local data is extracted to their full extent, in which case the models are defensible. This requires a statistical closure and more appropriate ways need to be developed. This paper just to demonstrate defensibility by visual means and counting the number of residual data points within a statistical range, although other approaches are also feasible.
The performance of any individual model is studied using Nash–Sutcliffe coefficient (NSC) and RMSE for the models at Levels 1 and 2 at calibration/training and validation/testing phases. For a ‘perfect’ model, the NSC value is close to 1; whereas RMSE is close to 0. Poor performances are reflected in NSC lower than 1 and RMSE higher than 0.

The literature review indicates that two types of performance metrics are popularly reported in statistical downscaling studies: (i) performance metrics for all datasets in the time scale of monthly or daily and (ii) performance metrics for averaged values in each of 12 months of years. The review shows that performance metrics of the first type are significantly less popular than the second type. Gulacha et al. (2016) listed a set of studies related to the statistical downscaling of precipitation and showed that $R^2$ varies in the range of 0.06–0.52 as per the first type of performance metrics. However, performance metrics as per the second type are higher than the first type, albeit this has not been explicitly stated in previous studies. The first type of performance metrics is reported in the study.

2.6. Data availability and study area

The study area, shown in Figure 2, comprises two grid cells, where the main cell is located at the south of the Caspian Sea and the north of Iran (see Figure 1) largely covering southern Gilan province and the second grid covers territories of East Azerbaijan, Ardabil and Zanjan provinces. These cells belong to Can-ESM2 (2.79° × 2.81° – lon° × lat°) with the resolution of 309 × 309 km², as used in the 5th Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC). The synoptic station is located in the main cell, but Grid Cell 2 is also considered as its climatic pattern can affect the climate of adjacent grid cell. Notably, Can-ESM2 data are available on the website of Canadian Climate Data and Scenarios (http://climate-scenarios.canada.ca). Table 2 presents the list of predictors from Can-ESM2.

The study uses the data obtained from the synoptic station of Rasht, Gilan province, run by the Iran Methodological Organisation. The annual precipitation is 1,100 mm in the study area, which is approximately 4× greater than the average precipitation in the country for that particular 10-year period (2004–2013). This amount of precipitation produces approximately $4,500 \times 10^6$ m³ runoff throughout the province (Gilan Regional Water Authority). The annual maximum and minimum temperatures in the same time period are 37.8 and −3.5 °C, respectively. According to Emberger (1930), the study area is characterised by cold and humid climate.

The study uses two sets of data: (i) predominant predictors among NCEP/NCAR reanalysis data are listed in Table 1 and (ii) the monthly precipitation for the synoptic station of Rasht serving as the predictand is the downscaling process. Predictors and the predictand were divided into 1961–1995 and 1996–2005 for the training phase (or calibration) and the testing phase (or validation), respectively. Accordingly, 540 data points were used in the training phase and 120 data points in the testing phase. Notably, training/testing is common terminologies in AI-based modelling, and calibration/validation is used for statistical (regression) modelling.

The study also uses a third set of results based on using GCM scenario results to project the calibrated and tested modelling results into the future and best practice procedure, which are outlined next. The paper uses Can-EMS2 to extract future climate variables under RCP2.6, RCP 4.5 and RCP8.5 scenarios. The difference between these scenarios lies in temperature changes during 2011–2040, 2041–2070 and 2071–2100 (Table 3).

### 3. MODELLING RESULTS

#### 3.1. Level 0: pre-processing and identification of predominant predictors

In the pre-processing stage, the data quality was studied by conducting the outlier test as included in Grubbs’s test on local precipitation data and analysing the gap data. No outlier data were identified and the number of data gaps was 5 months, which does not impair the research in terms of data quality. The pre-processing activities on data also comprise the selection of GCMs.

| RCP  | 2011–2040 | 2041–2070 | 2071–2100 |
|------|-----------|-----------|-----------|
| 2.6  | 0.75      | 0.78      | 0.88      |
| RCP4.5 | 1.07     | 1.44      | 2.07      |
| RCP8.5 | 1.06     | 1.8       | 3.55      |

*Representative concentration pathways.
grid cells around the synoptic station at a study area and that of predominant predictors. Notably, this stage may involve a selection from Can-ESM2, Had-CM3, BNU-ESM, HadGEM2-AO and CGCM3(T63), Can-ESM2 using CC between observed and predicted predictand but it is considered as out of the scope of the study. Two grid cells were selected after examining other grid cells adjacent to the grid cells in Figure 2. Based on preliminary evaluations, predominant predictors are selected within the grid cells and shown in Figure 2. This may refer to: (i) the Alborz mountain at the south of the study area which breaks climatic connection between the study area and grid cells at south and (ii) the Caspian Sea at north which affects the climatic predictors within corresponding grid cells and weakens its correlation with the cells shown in Figure 2.

The input datasets for models at Level 1 are predominant predictors screened from 26 NCEP/NCAR predictors belonging to Can-ESM2 at two grid cells (see Figure 2), which are selected by CC for SDSM and by decision tree for AI models. The result of predictors screening by CC indicates that four predictors at Grid Cell 1 are selected as the predominant predictors and they comprise (i) precipitation; (ii) meridional wind component at 850 hPa; (iii) wind speed at 1,000 hPa and (iv) meridional wind component at 1,000 hPa. Also, the result of predictors screening by decision tree indicates that six predictors are selected at both Grid Cells 1 and 2, which comprise (i) precipitation; (ii) mean sea level pressure at Grid Cell 1; (iii) divergence of true wind at 850 hPa; (iv) wind direction at 1,000 hPa; (v) geopotential height at 850 hPa and (vi) specific humidity at 850 hPa at Grid Cell 2. Notably, the predominant predictors identified by both approaches are not identical in terms of the type of predictors and the number of grid cells. The predictors are normalised between 0 and 1 to eliminate the effect of variation range.

3.2. Level 1: results

Figure 3(a)–3(c) presents precipitation time series predicted by SDSM, SVM and SFL at Level 1 for the training and testing phases and compares them against their observed precipitation time series. A visual comparison between the predicted and observed time series indicates that models at Level 1 are fit-for-purpose and confirms that the performance somewhat deteriorates from training to testing phases, as expected. In the testing period, predicted values seemingly tend to be higher values than those by AI models. Also, some of the predicted extreme values suffer from larger errors by all models during both training and testing phases, but even in this case, the behaviours of predicted time series are similar to the observed values.

Visual displays of the time series of predicted and observed values are limited in their scope, but a critical understanding emerges readily by a study of their scatter diagrams, as presented in Figure 4(a1), 4(b1) and 4(c1), displaying observed versus predicted precipitation for training and testing phases of SDSM, SVM and SFL at Level 1. A model with higher performance has lower dispersion against the 1:1 line. A more critical understanding emerges by presenting residual errors of models versus observed precipitation, as given by Figure 4(a2), 4(b2) and 4(c2), according to which a model with higher performance has lower dispersion against the line with zero residuals. Figure 4(a2) and 4(c2) indicates that SVM (4b2) has relatively lower dispersion than SFL (4c2) and SDMS (4a2), and SFL (4c2) has a lower dispersion than SDSM (b2). However, for the higher values of precipitation, there is a slight trend towards the upper part of the diagrams, and therefore extreme values of precipitation can be relatively underestimated.

A further study of the models at Level 1 is presented in Table 4 in terms of performance metrics of NSC and RMSE, although the intention is not to rank these models. As per NSC and RMSE, SVM performs better than SFL and SDSM, and SFL performs better than SDSM.

Figure 5(a) and 5(b) compares averaged values of estimated and observed precipitation in every 12 months of the years during calibration/training and validation/testing phases, respectively. The averaged values of precipitation in every 12 months of the years during the whole period of data availability display a pattern of behaviour, which may be used as a convenient representative of the observed local precipitation data and will be used as a visual comparison basis to study the behaviour of the modelling results. The figure indicates that models have better performance metrics in the calibration/training period than the validation/testing and this is expectable. The figure also provides evidence that SVM performs relatively better than the SDSM and SFL, except for May (see Figure 5(a)) and May–July (see Figure 5(b)). The fact that SVM does not perform better than other models in all months justifies the necessity of combining models at Level 1 by IMM strategies.

3.3. Level 2: results of NF models

The results of the IMM-NF modelling strategy are also provided in Table 4, according to which the improvements in performance metrics are significant with only one exception that the performance of SDSM is slightly better. Further comparisons are outlined below using the information presented in Figures 3–5. A comparison between the predicted precipitation at Level 2
in Figure 3(d) and that at Level 1 (Figure 3(a)–3(c)) suggests visually a possible improvement at Level 2 but this needs further attention. The improvements in performance metrics (Table 4) and the quality of model fits (Figure 5) are evident. For instance, the time series display residual errors in Figure 3(e), which show that lower residuals during both training and testing phases belong to IMM-NF at Level 2.

The scatter diagram for IMM-NF at Level 2 is presented in Figure 4(d1) and 4(d2), which makes it possible to visually compare the scatter diagrams for models at both levels. It readily shows that NF has relatively lower dispersion than models at Level 1 and this reveals that the enhanced performance metrics by IMM-NF at Level 2 also give rise to its quality improvements. Figure 4(e1) and 4(e2) highlights this issue and quantifies the dispersivity of the results by counting data points located in the transparent yellow band within ±50 mm. The figure indicates that 76% of the results by IMM-NF are located within the band, whereas those for SDSM, SFL and SVM are 63, 69 and 69%, respectively. According to the figure and Table 4, IMM-NF at Level 2 has significantly improved the performance of individual base models at Level 1 in terms of quantified dispersivity and performance metrics. Also, Figure 5 indicates that predicted precipitation by IMM-NF has a lower deviation with respect to observed precipitation than other models.

3.4. Projections into the future

The SDSMs by IMM-NF modelling are employed to project future precipitations under different scenarios, the procedure for which is outlined above. The results of different scenarios are compared for 12 months of a year during each projected period.
Figure 4 | Scatter and residual error diagrams for SDSM and AI models at Levels 1 and 2. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/wcc.2021.106.

Table 4 | Performance metrics for models at Levels 1 and 2 in calibration/training and validation/testing phases

| Performance metrics | SDSM (Level 1) | SVM (Level 1) | SFL (Level 1) | Average at Level 1 | NF (Level 2) |
|---------------------|---------------|---------------|---------------|--------------------|--------------|
| NSC (Training)      | 0.632         | 0.685         | 0.707         | 0.675              | 0.747        |
| NSC (Testing)       | 0.360         | 0.500         | 0.384         | 0.415              | 0.550        |
| NSC (Total)         | 0.585         | 0.653         | 0.651         | 0.630              | 0.713        |
| RMSE (Training)     | 57.9          | 53.1          | 51.2          | 54.1               | 47.6         |
| RMSE (Testing)      | 65.2          | 57.6          | 63.9          | 62.2               | 54.7         |
| RMSE (Total)        | 59.2          | 54.1          | 54.3          | 55.9               | 49.2         |

Model with lowest performance | Model with highest performance
(2011–2040; 2041–2070 and 2071–2100). Figure 6(a) compares the results of different scenarios during 2011–2040 and confirms insignificant differences between different scenarios. Similar behaviour is observed for other periods (2041–2070 and 2071–2100), but the results are not presented for brevity. Also, the figure represents observed and predicted precipitation during 1961–2005 and indicates that the projected precipitation will be reduced compared with the past precipitation (1961–2005) on some months (October–February).

The projected precipitation by the models at Levels 1 and 2 are displayed in Figure 6(b)–6(d) for each projected period. According to the figure, the pattern of variations during different months is similar except for SDSM, which overestimates the projected precipitation from July to October. This issue is attributable to the linear relationship between predictors and the predictand in SDSM, whereas AI models seek nonlinear relationships. Notably, significant differences are not observed between models in the downscaling step. The comparison between the projected precipitation with observed precipitation indicates that generally, the projected precipitation will be reduced with regard to the observed precipitation from 1961–2005.
January to February and October to December despite some reversals. Also, the projected precipitation will be slightly increased with regard to the observed precipitation from April to May.

4. DISCUSSION

The IMM strategy is in its infancy and the ongoing research activities of the authors demonstrate its feasibility in different fields of water resources engineering. This paper presents a new application of IMM in climate studies, in which IMM incorporates a set of models at Level 1 and combines their results at Level 2 in the downscaling and projection stages. The paper reveals an interesting feature of IMM, in which models have approximately identical behaviour at the downscaling stage during both the training and testing phases. Thus, there is not much difference between the modelling results produced at Levels 1 and 2 compared with the pattern of the averaged values of precipitation in every 12 months of the years as discussed in Section 3.2 and Figure 5. However, the modelling results at the projection stage display some discordant outcomes as displayed in Figure 6, in which the models at Level 1 are somewhat discordant in comparison with the above-mentioned pattern, whereas the projected modelling results at Level 2 are in concordance with the above pattern. The results presented in the paper serve as a proof-of-concept for the applications of IMM to modelling studies on climate change, as they enhance accuracy and provide new insights. However, there are no published results on similar applications to climate change, and therefore there are no comparable results to discuss in the paper.

The idea of statistical downscaling by multiple models is an ongoing research activity in the literature and categorised into (i) ‘ensemble models’ which performs downscaling by different AI or regression models and compares the results (e.g., Li et al. 2020) and (ii) using statistical models to combine the result of downscaling obtained by different models (e.g., Su et al. 2019). Although the second category can be included in the IMM modelling strategy, it is necessary to capture deeper information on the performance of AI models as a downscaling or combiner model. This study provides a proof-of-concept for using IMM in downscaling and projecting precipitation, in which the data are taken from one synoptic station. A consideration for several stations using an IMM strategy is planned by the authors to use the case of the Lake Urmia basin, where the lake is shrinking catastrophically.

The formulated IMM modelling strategy consists of two levels in which SDSM, SFL and SVM are trained and tested at Level 1 and their outputs are fed to NF at Level 2. Notably, the choice for selecting the models at both levels and the number of modelling strategies are wide. However, Khatibi et al. (2020) also discuss the possibility of analysing statistical distribution of error residuals to set a closure on minimising the number of strategies but these are outside the scope of the paper, although they are planned for future studies.

The monthly time scale for precipitation downscaling was investigated in the study, but other downscaling resolutions are practised: a higher resolution of 3 h was used by Mendes & Maia (2020); the daily interval was used by various authors (see Table 1) and lower resolutions of the monthly intervals (see Table 1) or annual interval by Sachindra & Perera (2018). Generally, higher resolutions are appropriate to investigate hydrological models such as rainfall–runoff (e.g., TavakoliFar et al. 2017), whereas lower resolutions are used for groundwater (e.g., Tewari et al. 2015) or drought studies (e.g., Tabari et al. 2020). Considering the importance of downscaling resolution, there is room for investigating the applicability of IMM for downscaling and projecting precipitation at different time resolutions.

5. CONCLUSION

This study formulated an IMM strategy to investigate the response of a study area to climate change through downscaling and projection of monthly precipitation. The IMM strategy incorporates three models at Level 1, which comprise SDSM, SFL and SVM, and feeds their outputs to NF at Level 2. The modelling activities also included the selection of predominant predictors among the NCEP/NCAR predictors using appropriate techniques. The results indicated that models at Level 1 are fit-for-purpose, but the model at Level 2 (NF) significantly improves the downscaling accuracy in terms of less dispersion in residual errors and performance metrics. All models somewhat underestimated the extreme precipitation values.

The future is uncertain and the paper presents the result, knowing that its projection for the next hundred-year will be highly uncertain; nonetheless, the paper puts an effort to improve the reliability of projected precipitation by the formulated IMM strategy in the downscaling stage. At a first glance, it was observed that although the models presented close results in the downscaling stage, one of the models at Level 1 (SDSM) showed a largely different behaviour during some months in the projection stage. IMM reduces this discordance between Level 1 models, which is not evident in the downscaling stage. Also,
the projection results indicate that precipitation will be reduced compared with observed precipitation in cold seasons (October–February), but the projected precipitation will be slightly increased in wet seasons (April to May).

The IMM strategy formulated by the study provided an insight into downscaling and projection of precipitation. However, further investigations may be suggested based on the limitations in the study summarised as follows: (i) an IMM modelling strategy was formulated, in which the aim is to learn both from local data, as well as from different GCM models; (ii) the residual analysis provided evidence that the accuracy is enhanced by increasing precipitation and there is room for further improvements by investigating different strategies to increase the performance; (iii) monthly precipitation was investigated by the paper but lower and higher resolutions for both precipitation and temperature can be considered and (iv) the performance of IMM was highlighted in the projection stage and this may be referred to the uncertainty associated between models, and therefore an uncertainty analysis can be carried out for a better insight.

**DATA AVAILABILITY STATEMENT**

All relevant data are available from an online repository or repositories.

**REFERENCES**

Ahmed, K., Sachindra, D. A., Shahid, S., Iqbal, Z., Nawaz, N. & Khan, N. 2020 Multi-model ensemble predictions of precipitation and temperature using machine learning algorithms. *Atmospheric Research* **236**, 104806.

Anandhi, A., Srinivas, V. V., Nanjundiah, R. S. & Nagesh Kumar, D. 2008 Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine. *International Journal of Climatology: A Journal of the Royal Meteorological Society* **28** (3), 401–420.

Chen, S. T., Yu, P. S. & Tang, Y. H. 2010 Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *Journal of Hydrology* **385**, 13–22.

Chiu, S. 1994 Fuzzy model identification based on cluster estimation. *Journal of Intelligent & Fuzzy Systems* **2**, 267–278.

Clemen, R. T. 1989 Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting* **5** (4), 559–583.

Ghorbani, M. A., Khatibi, R., Karimi, V., Yaseen, Z. M. & Zounemat-Kermani, M. 2018 Learning from multiple models using artificial intelligence to improve model prediction accuracies: application to river flows. *Water Resources Management* **32** (13), 4201–4215.

Gulacha, M. M. & Mulungu, D. M. 2017 Generation of climate change scenarios for precipitation and temperature at local scales using SDSM in Wami-Ruvu River Basin Tanzania. *Physics and Chemistry of the Earth, Parts A/B/C* **100**, 62–72.

Hammani, D., Lee, T. S., Ouarda, T. B. & Lee, J. 2012 Predictor selection for downscaling GCM data with LASSO. *Journal of Geophysical Research: Atmospheres* **117** (D17).

Hashemi, M. Z., Shamseldin, A. Y. & Melville, B. W. 2011 Statistical downscaling of watershed precipitation using gene expression programming (GEP). *Environmental Modelling & Software* **26**, 1639–1646.

Hosseini, R. H., Golian, S. & Yazdi, J. 2020 Evaluation of data-driven models to downscale rainfall parameters from global climate models outputs: the case study of Lybian watershed. *Journal of Water and Climate Change* **11** (1), 200–216.

Ishida, K., Tanaka, K. & Hama, T. 2020 Sensitivity analysis of convective parameterizations of a regional climate model in higher-resolution domains for long-term precipitation reconstruction. *Journal of Water and Climate Change* **11** (4), 1467–1480.

Jafarzadeh, A., Pourreza-Bilondi, M., Siuki, A. K. & Moghadam, J. R. 2021 Examination of various feature selection approaches for daily precipitation downscaling in different climates. *Water Resources Management* **35** (2), 407–427.

Jang, J. S. 1993 ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics* **23** (3), 665–685.

Jang, J. S. R., Sun, C. T. & Mizutani, E. 1997 Neuro-fuzzy and soft computing—a computational approach to learning and machine intelligence [book review]. *IEEE Transactions on Automatic Control* **42** (10), 1482–1484.

Karimi, V., Khatibi, R., Ghorbani, M. A., Bui, D. T. & Darbandi, S. 2020 Strategies for learning groundwater potential modelling indices under sparse data with supervised and unsupervised techniques. *Water Resources Management: An International Journal*, 1–29.

Khatibi, R., Ghorbani, M. A., Naghshehara, S., Aydin, H. & Karimi, V. 2020 Introducing a framework for inclusive multiple modelling with critical views on modelling practices – applications to modelling water levels of Caspian Sea and Lakes Urmia and Van. *Journal of Hydrology* **124923**.

Khatibi, R. & Nadiri, A. A. 2020 Inclusive multiple models (IMM) for predicting groundwater levels and treating heterogeneity. *Geoscience Frontiers*. Li, X., Li, Z., Huang, W. & Zhou, P. 2020 Performance of statistical and machine learning ensembles for daily temperature downscaling. *Theoretical and Applied Climatology*, 1–18.

Mendes, J. & Maia, R. 2020 Spatial downscaling of 3-hourly precipitation forecast data at river basin scale. *Meteorology and Atmospheric Physics* **132** (1), 143–158.

Moazamnia, M., Hassanzadeh, Y., Nadiri, A. A., Khatibi, R. & Sadeghfam, S. 2019 Formulating a strategy to combine artificial intelligence models using Bayesian model averaging to study a distressed aquifer with sparse data availability. *Journal of Hydrology* **571**, 765–781.

Moazamnia, M., Hassanzadeh, Y., Nadiri, A. A. & Sadeghfam, S. 2020 Vulnerability indexing to saltwater intrusion from models at two levels using artificial intelligence multiple model (AIMM). *Journal of Environmental Management* **255**, 109871.
Nadiri, A. A., Sedghi, Z., Khatibi, R. & Sadeghfam, S. 2018 Mapping specific vulnerability of multiple confined and unconfined aquifers by using artificial intelligence to learn from multiple DRASTIC frameworks. Journal of Environmental Management 227, 415–428.

Nadiri, A. A., Khatibi, R., Khalifi, P. & Feizizadeh, B. 2020 A study of subsidence hotspots by mapping vulnerability indices through innovative ALPRIFT using artificial intelligence at two levels. Bulletin of Engineering Geology and the Environment, 1–15.

Nourani, V., Razzaghzadeh, Z., Baghanam, A. H. & Molajou, A. 2019 ANN-based statistical downscaling of climatic parameters using decision tree predictor screening method. Theoretical and Applied Climatology 137, 1729–1746.

Pal, M. & Deswal, S. 2009 M5 model tree based modelling of reference evapotranspiration. Hydrological Processes: An International Journal 23 (10), 1437–1443.

Perevez, M. S. & Henebry, G. M. 2014 Projections of the Ganges–Brahmaputra precipitation – downscaled from GCM predictors. Journal of Hydrology 517, 120–134.

Quinlan, J. R. 1992 Learning with continuous classes. In: Proceedings of the Fifth Australian Joint Conference on Artificial Intelligence, Hobart, Australia, 16–18 November. World Scientific, Singapore, pp. 343–348.

Romero, E. & Toppo, D. 2007 Comparing support vector machines and feedforward neural networks with similar hidden-layer weights. IEEE Transactions on Neural Networks 18 (3), 959–963.

Sachindra, D. A., Ahmed, K., Rashid, M. M., Shahid, S. & Perera, B. J. C. 2018 Statistical downscaling of precipitation using machine learning techniques. Atmospheric Research 212, 240–258.

Sachindra, D. A. & Perera, B. J. C. 2018 Annual statistical downscaling of precipitation and evaporation and monthly disaggregation. Theoretical and Applied Climatology 131 (1), 181–200.

Sadeghfam, S., Ehsanitabar, A., Khatibi, R. & Daneshfaraz, R. 2018 Investigating ‘risk’ of groundwater drought occurrences by using reliability analysis. Ecological Indicators 94, 170–184.

Sadeghfam, S., Daneshfaraz, R., Khatibi, R. & Minaei, O. 2019 Experimental studies on scour of supercritical flow jets in upstream of screens and modelling scouring dimensions using artificial intelligence to combine multiple models (AIMM). Journal of Hydroinformatics 21 (5), 893–907.

Sedwick, P. 2012 Pearson’s correlation coefficient. The BMJ 345, e4483.

Semenov, M. A. & Barrow, E. M. 1997 Use of a stochastic weather generator in the development of climate change scenarios. Climatic Change 35 (4), 397–414.

Su, H., Xiong, Z., Yan, X. & Dai, X. 2019 An evaluation of two statistical downscaling models for downsampling monthly precipitation in the Heihe River basin of China. Theoretical and Applied Climatology 138 (3–4), 1913–1923.

Teegavarapu, R. S. & Goly, A. 2018 Optimal selection of predictor variables in statistical downscaling models of precipitation. Water Resources Management 32 (6), 1969–1992.

Tewari, R., Johnson, J., Mauger, S., Leiker, G., Hayhoe, K., Hernandez, A., Rainwater, K. et al. 2015 Using climate scenarios to evaluate future impacts on the groundwater resources and agricultural economy of the Texas High Plains. Journal of Water and Climate Change 6 (3), 561–577.

Vapnik, V. N. 1998 An overview of statistical learning theory. IEEE Transactions on Neural Networks.