JOURNAL OF
HYDRO-METEOROLOGY

Check for updates

Research Article

Evaluate the correct and the skill of the IFS model for minimum temperature, average temperature, maximum temperature forecasting in short term (24 hours) at 09 regions in Vietnam

Le Thi Thu Ha[1]*, Nguyen Thu Hang[2], Tran Thi Thanh Hai[1], Nguyen Thi Tuyet[3]

[1] Meteorological and Hydrological Forecasting Management Department; leha246@gmail.com; haitran84@gmail.com
[2] National Center for Hydro-Meteorological Forecasting; nthang0676@gmail.com
[3] Ho Chi Minh University of Natural Resource and Environment; nttuyet@hcmunre.edu.vn

*Corresponding author: leha246@gmail.com; Tel.: +84–904290269

Received: 12 February 2024; Accepted: 18 March 2024; Published: 25 March 2024

Abstract: Conceptually, forecast verification is simple, you just need to compare the forecast factors and observed factors. The accuracy of a forecast is a measure of how close to the actual weather the forecast was. The reliability of a forecast is the average agreement between the forecast values and the observed values. The skill of a forecast is performed based on some benchmark forecast, usually by comparing the accuracy of the forecast with the accuracy of the benchmark. The benchmark forecast can be a climatic value. Meanwhile, the correct forecast is bias between the forecast value and the observed value within the allowable range. This study evaluates the correct and forecasting skill of the IFS model (by European Centre for Medium-Range Weather Forecasts) for minimum temperature (T_m), average temperature (T_{ave}), maximum temperature (T_x) forecasting in 24 hours at 09 regions in Viet Nam. The results show that within 24 hours, the IFS model predicts a high bias for the T_m (from 0.2 to 0.9^{o}C) and a low bias for the T_{ave} (from -0.2 to -0.9^{o}C) and T_x (from -1.0 to -2.0^{o}C). The correct in the southern region is higher than in the northern region (average about 10 to 15%). The skill of IFS model is higher than the benchmark (skill for the T_m has exceeded the Benchmark value by 0.4 to 0.6; skill for the T_{ave} has exceeded the Benchmark value by 0.5 to 08), in there, the skill of T_m and T_{ave} is higher than skill of T_x at the most regions, except in the Southern region, the skill of IFS model is lower than the benchmark for T_{ave} and T_x.

Keywords: Accuracy; Reliability; Skills; Forecast Verification.

1. Introduction

According to Guidelines of World Meteorological Organization (WMO) [1], the general purpose of the verification is to ensure that the forecast and warning products are accurate, competent, and reliable from a technical point of view. This is distinct from whether the products are actually meeting user needs. However, technical verification must be based on methods appropriate to the user's needs. There are many studies on verification methods. Allan Murphy, a pioneer in the field of forecast verification, wrote an essay on what makes a forecast "good" [2], a good forecast is a forecast that satisfies the following three criteria: Consistency: the level of forecast changes according to changes in situation; Good quality: the degree of agreement between forecast and observation; Valuable: the extent to which the forecast supports decision making and brings benefits; Also according to the research of [2], forecast quality includes the following nine attributes: Bias;

Correlation; Accuracy; Forecasting skill; Reliability; Resolution; Sharpness; Discrimination and Uncertainty. Simply, forecast verification includes accuracy and skill. Note that, the other attributes of forecast quality also affect the value of the forecast.

According to the research of [3–5] describe methods for assessing the value of the forecasts. Forecast quality is not the same as forecast value. High forecast quality if the forecast and observation are well according to some objective criteria. Forecast value helps the user to make a better decision.

Meanwhile, regarding the verification results, according to the research of [6–9]: the verification results are more reliable when the quantity and quality of verification data are high. The usual approach is to determine the confidence interval for the verification score using approximate, analysis methods; Regarding stratification results, to obtain reliable verification statistics, the verification data should be divided by time and space. For example, according to the study [10], the verification data is divided by season, geographical region, monitoring frequency, etc.

Regarding the standard verification methods, there are many studies for greater detail of the standard verification methods see [11] or one of the excellent the research of [12], [13–15] on forecast verification and statistics. The results see that, with methods for forecast of continuous variables such as temperature, the verification indices as Bias, MAE, MSE and RMSE. These verification indices are simple and useful to explain to users before making decisions. The Bias index indicates the direction of the forecast bias, where the MAE and RMSE indices indicate the average amplitude of the forecast error. Therefore, people often use a combination of these indicators to provide an estimate of reliability.

In Viet Nam, Meteorological and Hydrological Administration has been invested in by the Ministry of Natural Resources and Environment to buy products (images are available on the page website: http://www.ecmwf.int) and numerical data (GRIB code transmitted over the Internet) of the European Centre for Medium-Range Weather Forecasts (ECMWF) to serve operational forecasting since the end of 2011. The data source of the ECMWF is considered plentiful with high reliability. Besides, some studies related to assessing the skills of models, including the IFS model such as studies by [16–18]. The studies mainly evaluated the skill for rainfall forecast and show that: Both skill validations of station-based and spatial-based show low skills of models for high thresholds of 24h accumulated rainfall forecast [18]. The IFS model has best forecast skill in comparison with the other models. However, all given model is under-estimating in forecasting extreme heavy rainfall events [16]; For rainfall quantity forecast, IFS model has skill from 24 hours to 48 hours lead time and less skill at 72 hours lead time. However, IFS model has skill for number of heavy rainfalls [17]. As for temperature, research by [19] shows: With the using of automatic calibration method, the forecast quality of the IFS model is significantly improved. According to the provisions of legal documents on verification the quality of hydro-meteorological forecasting and warning [20], the reliability is understood as determining the allowable error between the forecast value and the observed value. If the forecast value is within the allowable error range that mean correct; if it is outside the allowable error range, that mean not correct. Accordingly, the correct of forecast value is determined within ± 1 level compared to the observed value; Long term will have a wider allowable error range than short term. For meteorological natural disasters such as tropical storms, heavy rainfall, and heat waves, in addition to evaluating the forecast value, also evaluate the time of influence and scope of influence. According to legal documents is also assessed through the “completeness” of newsletter content and “timeliness” of newsletter delivery. For the temperature, within a forecast period from 1 to 3 days, allowable error ranges from -2^{o}C to 2^{o}C.

In general, there are not many detailed studies in Vietnam for temperature forecast factors, most of the studies focus on the standard verification methods, with few studies related to regulations at the legal documents.

This study will initially evaluate the skill of the IFS model for temperature forecasting in short term at all regions in Viet Nam. In addition, to determine whether the model can be applied in operational forecasting, the study will also evaluate the correct of the model following the legal documents.

2. Materials and Methods

2.1. Description of study site

Figure 1 presents a map of the study area, there are 09 regions in Viet Nam: 1) the Northwestern region; 2) the Mid-North region; 3) the North Eastern region; 4) the Red River Dela region; 5) the North Central region; 6) the Mid-Central region; 7) the South Central region; 8) the Central Highland region; 9) the Southern region.

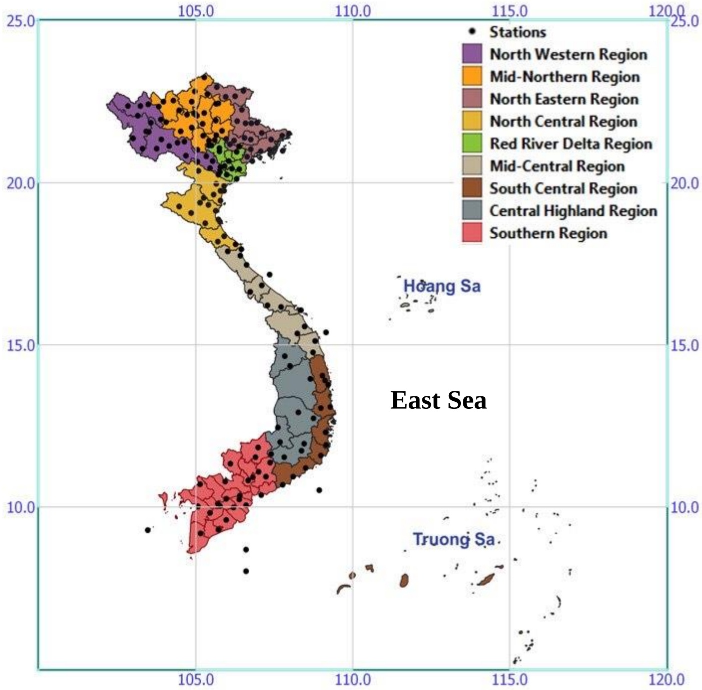

Figure 1. The study area at 09 regions in Viet Nam.

2.2. Data collection

In this study used: Observed data of daily minimum temperature, average temperature, maximum temperature from December 2019 to December 2022 of 184 synoptic stations in Viet Nam and shown in Table 1.

Climatic data from 1981 to 2010 of minimum temperature, average temperature, maximum temperature of 138 synoptic stations in Viet Nam (Table 1).

Table 1. Information about synoptic stations at 9 regions in Viet Nam.

Rg	Code	Station	Rg	Code	Station	Rg	Code	Station	Rg	Code	Station
Northwestern	48/01	Muong Te	Northeastern	48814	Vinh Yen	North Central	48840	Thanh Hoa	Central Highland	48865	Kon Tum
	48/02	Sin Ho		48/52	Tam Dao		48/70	Nhu Xuan		48866	Playcu
	48/03	Tam Duong		48808	Cao Bang		48/72	Tinh Gia		48867	An Khe
	48/06	Than Uyen		48/33	Bao Lac		48/74	Quy Chau		48868	Yaly
	48800	Muong Lay		48/40	Ng. Binh		48844	T. Duong		48872	Ayunpa
	48/09	Tuan Giao		48/43	T. Khanh		48/75	Quy Hop		48876	EaHleo
	48/10	Pha Din		48807	That Khe		48/76	Tay Hieu		48878	Buon Ho
	48811	Dien Bien		48830	Lang Son		48/79	Con Cuong		48/98	M Drak

Rg	Code	Station	Rg	Code	Station	Rg	Code	Station	Rg	Code	Station
	48/07	Phieng Lanh		48/46	Mau Son		48/77	Quynh Luu		48875	B.M. Thuot
	48/05	Muong La		48/47	Bac Son		48/80	Do Luong		48869	EaKmat
	48806	Son La		48/48	Huu Lung		48/81	Hon Ngu		48885	Lak
	48/16	Song Ma		48/49	Dinh Lap		48845	Vinh		48882	Dac Mil
	48/17	Co Noi		48838	Mong Cai		48/82	Huong Son		48886	Dak Nong
	48/18	Yen Chau		48/50	Quang Ha		48846	Ha Tinh		48880	Da Lat
	48/19	Bac Yen		48837	Tien Yen		48/84	Huong Khe		48881	Lien Khuong
	48/20	Phu Yen		48834	Co To		48/73	Hoanh Son		48884	Bao Loc
	48/25	Moc Chau		48836	Cua Ong		48/86	Ky Anh		48/83	Cat Tien
	48/26	Mai Chau		48833	Bai Chay		48/87	Tuyen Hoa		48883	Phuoc Long
	48/61	Kim Boi		48/60	Uong Bi		48848	Dong Hoi		48895	Dong Phu
	48/63	Chi Ne		48/53	Hiep Hoa		48847	Ba Don		48898	Tay Ninh
	48/64	Lac Son		48/55	Luc Ngan		48/89	Con Co		48/78	Tri An
	48818	Hoa Binh		48/56	Son Dong		48849	Dong Ha		48896	Bien Hoa
	48803	Lao Cai		48809	Bac Giang		48/90	Khe Sanh		48/71	Ta Lai
	48/30	Bac Ha		48/54	Bac Ninh		48852	Hue		48/88	Long Khanh
	48802	Sa Pa		48826	Phu Lien	Mid-Central	48/91	A Luoi		48899	Thu Dau Mot
	48/29	Pho Rang		48828	Hon Dau		48/92	Nam Dong		48894	Nha Be
	48/08	Mu.C.Chai		48839	Bach. L.Vi		48855	Da Nang		48903	Vung Tau
	48815	Yen Bai		48/57	Ba Vi		48/93	Tam Ky		48918	Con Dao
	48/14	Van Chan		48817	Son Tay		48/94	Tra My		48919	Huyen Tran
	48/35	Luc Yen		48820	Lang		48/85	Ly Son		48906	Moc Hoa
	48805	Ha Giang		48819	Hoai Duc		48863	Q.Ngai		48912	My Tho
	48/31	Hoang S Phi		48825	Ha Dong		48/95	Ba To		48911	Vinh Long
	48/32	Bac Me		48/59	Chi Linh		48/96	Hoai Nhon	Southern	48901	Ben Tre
Mid-Northern	48/34	Bac Quang		48827	Hai Duong		48864	An Nhon		48902	Ba Tri
	48/38	Dong Van	Red River Delta	48822	Hung Yen		48870	Quy Nhon		48908	Cao Lanh
	48812	T.Quang		48823	Nam Dinh		48/97	Son Hoa		48904	Cang Long
	48/36	Ham Yen		48829	Van Ly		48873	Tuy Hoa		48909	Chau Doc
	48/37	Chiem Hoa		48821	Phu Ly		48877	Nha Trang		48897	Tra Noc
	48/39	Cho Ra		48832	Nho Quan	South Central	48879	Cam Ranh		48910	Can Tho
	48/42	Ngan Son		48824	Ninh Binh		48892	Song T.Tay		48905	Vi Thanh
	48810	Bac Can		48/65	C.Phuong		48890	Phan Rang		48913	Soc Trang
	48831	Thai Nguyen		48835	Thai Binh		48887	Phan Thiet		48907	Rach Gia
	48/44	Dinh Hoa		48842	Hoi Xuan		48888	La Gi		48917	Phu Quoc
	48/23	Minh Dai		48/67	Yen Dinh		48889	Phu Quy		48916	Tho Chu
	48/51	Phu Ho		48/68	Sam Son		48891	Phan Ri		48915	Bac Lieu
	48813	Viet Tri		48/69	Bai Thuong		48861	Dak To		48914	Ca Mau

Forecast data of IFS model with information and shown in Table 2.

Table 2. Information about IFS model.

Resolution	Lead Time	Time Series	Products
0.125°	24 hours to 240 hours	2019 - 2022	Surface: Temperature 2m, Sea Level Pressure, rainfall, wind 10m (For the T_{ave}, calculated through the Temperature 2m, averaging the time periods 00, 06, 12 and 18z. The T_m and T_x are taken as the minimum value and maximum value during the period 00, 06, 12 and 18z of that day) Upper level: Geopotential Height, wind, Relative Vorticity, 1000-500mb (thickness&mslp), 300/200mb (divergence&wind)…

2.2. Methods

Figure 2 presents conceptual framework of the applied methodology in this study, in which the input data are from the IFS model, monitoring data of the stations, climatic data of the stations; Next, we process these data, statistic matrices and the results are verification indices.

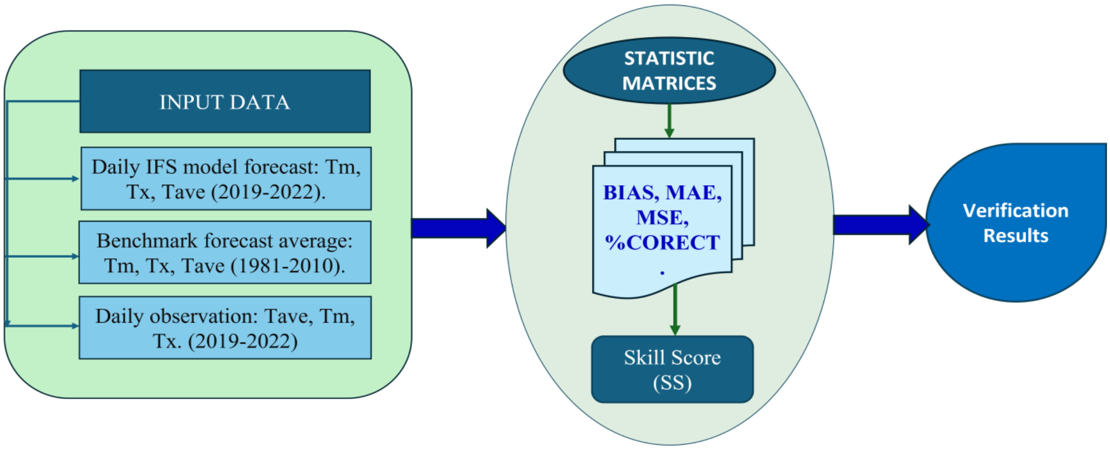

Figure 2. Conceptual framework of the applied methodology in this study.

There are many scientific documents on methods to forecast verification, for example the research of [3, 10]. According to WMO [1], there are two basic variables for forecasting: continuous variables (variables with numeric values) and grouped variables such as rain or no rain or hierarchical by intensity (light rain, moderate rain and heavy rain...). These variables can be predicted by giving specific values or by probabilities. Probabilistic forecasting will be more meaningful than numerical forecasting, in that users can make decisions based on probability and their perception.

The following simple example of a set of twenty maximum temperature forecasts will be used in this section to illustrate the score and shown in Table 3.

Table 3. Example for forecast verification indices.

MAX TEMP (ºC)					
Forecast (F)	**Observed (O)**	**F-O**	**ABS(F-O)**	**$(F-O)^2$**	**Within ± 2ºC**
17	17	0	0	0	1
24	20	4	4	16	0
28	29	-1	1	1	1
22	25	-3	3	9	0
14	16	-2	2	4	1
16	17	-1	1	1	1
17	17	0	0	0	1
16	16	0	0	0	1
15	14	1	1	1	1
19	18	1	1	1	1
22	19	3	3	9	0
21	17	4	4	16	0
16	18	-2	2	4	1
20	18	2	2	4	1
27	30	-4	4	16	0
21	20	1	1	1	1
15	14	1	1	1	1

MAX TEMP (ºC)						
	Forecast (F)	**Observed (O)**	**F-O**	**ABS(F-O)**	**(F-O)^2**	**Within ± 2ºC**
	22	28	-6	6	36	0
	20	23	-3	3	9	0
	15	18	-3	3	9	0
Average:	19.4	19.8	-0.4	2.1	6.9	60%
			Bias	MAE	MSE	% correct

a) Reliability

Suppose there are N forecasts f_i and corresponding observations o_i for i = 1...N

A gross measure of reliability is the mean bias. It is simply the average of the forecast value minus the average observed value as in equation (1).

$$\text{Bias} = \frac{1}{N}\sum_{i=1}^{N}\left(f_i - O_i\right) \tag{1}$$

For the example in Table 3, N=20, the average predicted value is 19.4°C and the average observed value is 19.8°C, so the average error value is -0,4°C, which means the forecast value is lower than the actual value. This is a simple method to determine reliability.

b) Accuracy

Various accuracy measures are shown in the previous table for this example. In terms of accuracy, the Mean Absolute Error or MAE is in equation (2):

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}(|f_i - o_i|) \tag{2}$$

The Mean-Square Error or MSE is presented in equation (3) and The Root-Mean-Square Error or RMSE is presented in equation (4).

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2 \tag{3}$$

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2} \tag{4}$$

According to the above example, a mean absolute error of 2.1°C means that the precision between the mean difference of predicted and observed temperature values is 2.1°C.

However, users are often interested in the largest possible error of the forecast, so will use the formula to calculate RMSE, which will be 2.6°C.

Another measure that is commonly used for weather elements such as temperature, is the "percent correct" of forecasts that are within some allowable range, e.g., within ±2°C or ±3°C. This is shown in the above table by putting a 1 when the forecast was within ±2°C of the observed maximum, and 0 otherwise, then averaging the values. The result for this example is that 60% of the forecasts are within ±2°C.

c) Skill

The skill of a forecast is exercised against some benchmark forecast, usually by comparing the accuracy of the forecast with the accuracy of the benchmark forecast. The benchmark forecast can be a climatic value or a value from an automated product.

For example, the climatic value of temperature during this period is 20°C, accordingly Table 4 gives the following evaluation results compared to the climatic value:

Table 4. Example for benchmark forecast verification indices.

MAX TEMP (ºC)						
	Benchmark Forecast (F)	**Observed (O)**	**F-O**	**ABS(F-O)**	**(F-O)^2**	**Within ± 2ºC**
	20	17	3	3	9	0
	20	20	0	0	0	1
	20	29	-9	9	81	0
	20	25	-5	5	25	0
	20	16	4	4	16	0
	20	17	3	3	9	0
	20	17	3	3	9	0
	20	16	4	4	16	0
	20	14	6	6	36	0
	20	18	2	2	4	1
	20	19	1	1	1	1
	20	17	3	3	9	0
	20	18	2	2	4	1
	20	18	2	2	4	1
	20	31	-11	11	121	0
	20	20	0	0	0	1
	20	14	6	6	36	0
	20	28	-8	8	64	0
	20	23	-3	3	9	0
	20	18	2	2	4	1
Average:	20.0	19.8	0.3	3.9	22.9	35%
			Bias	MAE	MSE	% correct

MAE_f is the absolute error for the forecast and MAE_b is the absolute error for the benchmark, then the forecast skill is calculated as (5):

$$1-\frac{MAE_f}{MAE_b}=1-\frac{2.1}{3.9}=0.45 \quad (5)$$

Or it can be calculated through the mean squared error of the forecast and the benchmark as (6):

$$1-\frac{MSE_f}{MSE_b}=1-\frac{6.9}{22.9}=0.7 \quad (6)$$

If the accuracy measure being used is the percent correct (of forecasts that are within an acceptable range of the observations), then another skill measure is as (7):

$$\frac{PC_f-PC_b}{100\%-PC_b}=0.38 \quad (7)$$

where the value of 0.38 means that the percent correct for the actual forecasts has gone 0.38 of the distance between the benchmark value of 35% and a perfect score of 100%.

d) Interpolation method

The grid data predicted from the model are interpolated to 184 synoptic station points using the bilinear interpolation method.

e) Regulation about the correct

Table 5 shows the correct used for forecast temperature in Clause 3, Article 12 of Circular No. 41/2017/TT-BTNMT of the Minister of Natural Resources and Environment promulgating technical regulations on assessing the quality of meteorological forecasting.

Table 5. The correct of forecast temperature according to regulation.

Error between forecast value and observed value	**Forecast time from 1 - 3 days**			**Forecast time from 4 - 10 days**		
	< -2ºC	**- 2ºC÷2ºC**	**> 2ºC**	**< -3ºC**	**- 3ºC÷3ºC**	**> 3ºC**
The reliability	-	+	-	-	+	-

3. Results and discussion

3.1. The correct and the skill of minimum temperature

Using WMO's guidelines, we calculated the BIAS, MAE, MSE, % correct indices of the IFS model from 2019-2022 for the T_m in 24 hours for 184 synoptic stations nationwide, then averaged at 09 regions in Viet Nam, the results are given in Table 6. Table 6 shows that for the 24-hour forecast period, the forecast T_m tends to be higher than the actual temperature from 0.2 to 0.9^{o}C in the most regions, except in the Northwestern region and the Red River Delta region, the forecast T_m tends to be lower than the actual temperature from -0.1 to -0.5^{o}C. The average amplitude of forecast error is largest in the Red River Delta region and smallest in the Southern region with MAE equals 1.4 and MSE equals 3.6^{o}C. With an allowed error range of ± 2^{o}C, the correct reaches from 96 to 98% for the southern provinces such as the Central Highland region, South Central region, and Southern region; the Mid-Northern region, Mid-Central region, and Northwestern region have the correct of 84 to 88%; the Red River Delta region, North Central region, and Northeastern region with the correct of 76 to 82%. This result is quite consistent because the monsoon circulation regime affecting the northern regions is more complex than the southern region, so the temperature variation in the northern regions is higher than the southern regions.

Table 6. The BIAS, MAE, MSE, % correct indices of the IFS model from 2019-2022 for the T_m in 24 hours at 09 regions in Viet Nam.

Region	Forecast (F)	Observed (O)	BIAS	MAE	MSE	%Correct
Northwestern	Average: 19.3	19.3	-0.1	1.0	1.8	88%
Mid-Northern	Average: 20.0	19.8	0.2	1.1	2.1	84%
Northeastern	Average: 20.6	20.4	0.2	1.2	2.6	82%
Red River Delta	Average: 21.4	21.9	-0.5	1.4	3.6	76%
North Central	Average: 22.3	22.2	0.2	1.2	2.7	80%
Mid-Central	Average: 23.7	22.8	0.9	1.1	1.9	86%
South Central	Average: 25.3	24.8	0.5	0.7	0.7	98%
Central Highland	Average: 20.4	20.1	0.2	0.7	0.8	96%
Southern	Average: 24.7	24.7	0.0	0.6	0.6	98%

The BIAS, MAE, MSE, % correct indices between the 30-year climatic average value (Benchmark) and the observed value from 2019-2022 for the T_m in 24 hours period are calculated for 138 synoptic stations, then averaged at 09 regions in Viet Nam, the results are given in Table 7.

Table 7. The BIAS, MAE, MSE, % correct indices of Benchmark from 2019-2022 for the T_m in 24 hours at 09 regions in Viet Nam.

Region	Benchmark (F)	Observed (O)	BIAS	MAE	MSE	%Correct
Northwestern	Average: 18.6	19.3	-0.7	1.8	5.3	66%
Mid-Northern	Average: 19.3	19.8	-0.5	1.9	6.1	64%
Northeastern	Average: 20.0	20.4	-0.4	2.0	7.0	58%
Red River Delta	Average: 21.1	21.9	-0.8	2.1	7.2	56%
North Central	Average: 21.1	22.2	-1.1	2.1	6.7	55%
Mid-Central	Average: 22.0	22.8	-0.7	1.4	2.9	77%
South Central	Average: 24.5	24.8	-0.3	0.7	0.9	96%
Central Highland	Average: 19.4	20.1	-0.8	1.2	2.1	83%
Southern	Average: 24.3	24.7	-0.4	0.8	1.0	96%

The skill of IFS model for T_m through comparison between the accuracy of the model's forecast value and the Benchmark is shown in Figure 3. In most regions, the model's forecast skill for the T_m has exceeded the Benchmark value by 0.4 to 0.6, especially in the Central Highland region, the percent correct for the actual forecasts has gone 0.8 of the distance between the benchmark value of 76% and a perfect score of 100%. However, the benchmark’s MAE and MSE are larger than the model’s MAE and MSE, that mean the benchmark's error is larger than the model’s error.

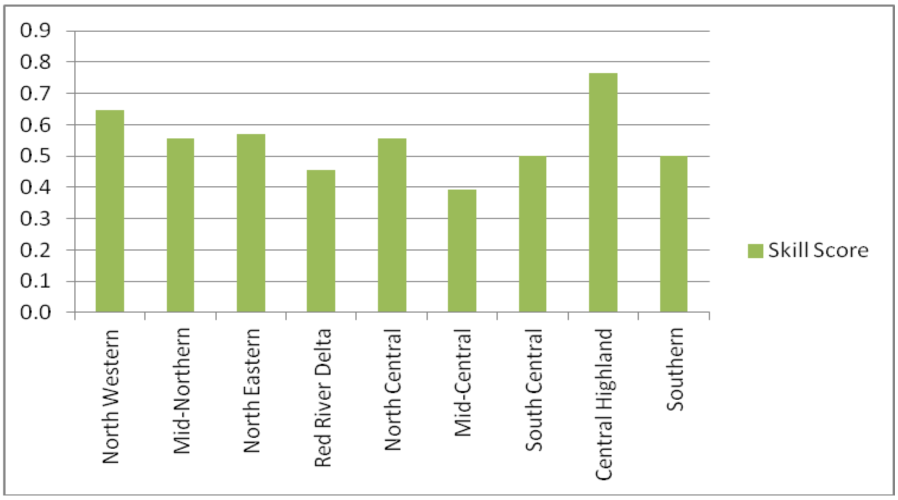

Figure 3. Skill score of IFS model for T_m at 09 regions in Viet Nam.

3.2. The correct and the skill of average temperature

The BIAS, MAE, MSE, % Correct indices of the IFS model from 2019-2022 for the T_{ave} in 24 hours for 184 synoptic stations nationwide, then averaged at 09 regions in Viet Nam, the results are given in Table 8.

Table 8. The BIAS, MAE, MSE, % correct indices of the IFS model from 2019-2022 for the T_{ave} in 24 hours at 09 regions in Viet Nam.

Region	Forecast (F)	Observed (O)	BIAS	MAE	MSE	%Correct
Northwestern	Average: 21.8	22.4	-0.6	1.1	1.7	89%
Mid-Northern	Average: 22.1	22.4	-0.3	1.0	1.6	91%
Northeastern	Average: 22.6	22.9	-0.3	1.0	1.6	91%
Red River Delta	Average: 23.5	24.3	-0.8	1.4	3.0	75%
North Central	Average: 24.3	24.5	-0.2	1.1	1.9	87%
Mid-Central	Average: 25.4	25.3	0.1	0.7	0.9	96%
South Central	Average: 27.2	27.2	-0.1	0.4	0.3	100%
Central Highland	Average: 22.9	23.5	-0.6	0.8	0.8	99%
Southern	Average: 26.6	27.5	-0.9	1.0	1.3	95%

Table 8 shows that for the 24-hour forecast period, the forecast T_{ave} tends to be lower than the actual temperature in most regions. The average amplitude of forecast error is largest in the Red River Delta region and smallest in the South-Central region, with MAE from 0.4 to 1.4^oC. With an allowed error range of ± 2^oC, the correct reaches from 95 to 100% in the southern provinces such as the Mid-Central region, South Central region, Central Highland region, and Southern region; the Northwestern region, the Mid-Northern region, Northeastern region, and North Central region have the correct of 87 to 91%; the Red River Delta region has the lowest correct of 75%.

Table 9 about the results of Benchmark’s BIAS, MAE, MSE, % Correct indices from 2019-2022 for T_{ave} at 09 regions shows that the correct is lower in the northern provinces and very high in the southern provinces, especially the correct in the South-Central region and Southern region reaches from 97 to 98%.

Table 9. The BIAS, MAE, MSE, % correct indices of Benchmark from 2019-2022 for the T_{ave} in 24 hours at 09 regions in Viet Nam.

Region	Benchmark (F)	Observed (O)	BIAS	MAE	MSE	%Correct
Northwestern	Average: 21.8	22.4	-0.6	1.8	5.4	62%
Mid-Northern	Average: 22.1	22.4	-0.3	1.9	5.9	61%
Northeastern	Average: 22.7	22.9	-0.2	2.0	6.4	59%
Red River Delta	Average: 23.5	24.3	-0.8	2.2	7.2	53%
North Central	Average: 23.8	24.5	-0.7	2.0	6.4	56%
Mid-Central	Average: 24.8	25.3	-0.5	1.5	3.6	71%
South Central	Average: 27.0	27.2	-0.2	0.7	0.8	98%
Central Highland	Average: 22.8	23.5	-0.7	1.1	1.8	88%
Southern	Average: 27.1	27.5	-0.4	0.8	1.0	97%

The skill of IFS model for T_{ave} through comparison between the accuracy of the model's forecast value and the Benchmark is shown in Figure 4. At the most regions, the model's forecast skill for the Tave has exceeded the Benchmark value by 0.5 to 08, especially in the Mid-Central region and the South Central region, the percent correct for the actual forecasts has gone 09 to 1. Except in the Southern region, the model's forecast skill for the Tave has lower than the Benchmark. This can be explained by the fact that the temperature regime in the Southern region has little change. Climatic values can be used to predict the average temperature of this region. That like the T_m, the benchmark's MAE and MSE are larger than the model's MAE and MSE for T_{ave}.

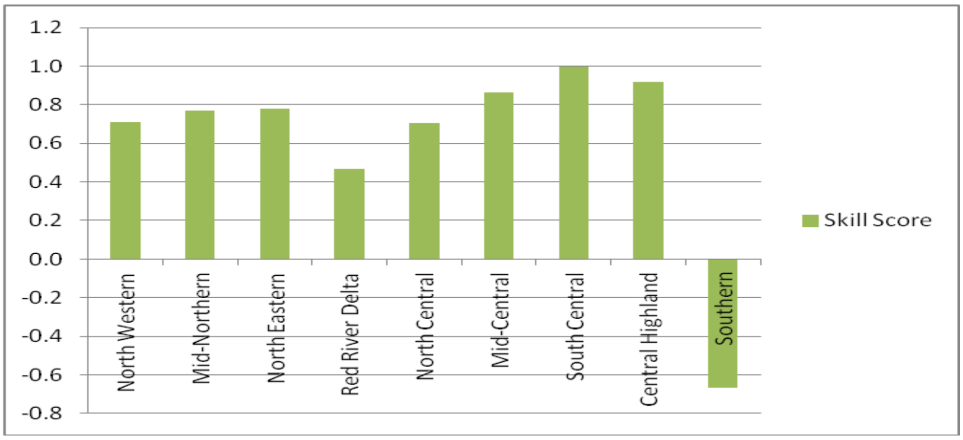

Figure 4. Skill score of IFS model for T_{ave} at 09 regions in Viet Nam.

3.3. The correct and the skill of maximum temperature

With the same method, the results of the BIAS, MAE, MSE, % Correct indices of the IFS model from 2019-2022 for the T_x within 24-hour period at 09 regions in Viet Nam is given by Table 10.

Table 10. The BIAS, MAE, MSE, % correct indices of the IFS model from 2019-2022 for the T_x in 24 hours at 09 regions in Viet Nam.

Region	Forecast (F)	Observed (O)	BIAS	MAE	MSE	%Correct
Northwestern	Average: 26.2	27.7	-1.5	2.4	7.9	45%
Mid-Northern	Average: 25.7	26.7	-1.0	2.1	6.5	53%
Northeastern	Average: 25.7	26.9	-1.2	2.1	6.5	50%
Red River Delta	Average: 26.9	28.0	-1.0	2.3	7.8	50%
North Central	Average: 27.7	28.4	-0.7	1.9	5.8	61%
Mid-Central	Average: 28.4	29.4	-1.0	1.7	4.0	66%
South Central	Average: 30.6	31.1	-0.5	0.9	1.3	91%
Central Highland	Average: 27.9	28.9	-1.1	1.5	3.3	72%
Southern	Average: 30.1	32.1	-2.0	2.1	5.6	51%

Table 10 shows that for the 24-hour forecast period, the forecast T_x tends to be lower than the actual temperature at the most regions, about from 1.5 to 2.0^oC. The average amplitude of forecast error is largest in the Northwestern region and the Red River Delta region and smallest in the South-Central region (MAE is approximately 2.4^oC, MSE is approximately 8^oC in the northern regions, meanwhile MAE and MSE are only approximately 1^oC in the South-Central region). With an allowed error range of ± 2^oC, the South-Central region has the highest correct of 91%; the North Central region, the Mid-Central region, and the Central Highland region have the correct of 61 to 72%; the other regions have the lower correct of 45 to 53%.

Table 11. The BIAS, MAE, MSE, % correct indices of Benchmark from 2019-2022 for the T_x in 24 hours at 09 regions in Viet Nam.

Region	Benchmark (F)	Observed (O)	BIAS	MAE	MSE	%Correct
Northwestern	Average: 27.4	27.7	-0.3	2.7	11.3	43%
Mid-Northern	Average: 26.6	26.7	-0.1	2.5	9.8	47%
Northeastern	Average: 27.0	26.9	0.1	2.5	9.8	48%
Red River Delta	Average: 27.1	28.0	-0.9	2.8	11.4	38%
North Central	Average: 28.0	28.4	-0.5	2.6	10.4	44%
Mid-Central	Average: 29.1	29.4	-0.3	2.1	7.5	55%
South Central	Average: 30.7	31.1	-0.4	1.0	1.6	89%
Central Highland	Average: 28.4	28.9	-0.5	1.6	3.8	66%
Southern	Average: 31.5	32.1	-0.6	1.2	2.0	85%

Table 11 shows that, for the T_x, Benchmark has quite high correct in the South-Central region and the Southern region with reaches from 85 to 89%. In the remaining regions, the correct is lower, only about 40 to 60%.

The skill of IFS model for T_x is shown in Figure 5, which shows that, skill of T_x is lower than skill of T_m and T_{ave}, only exceeding 0.1 to 0.2 compared to the benchmark value at the most regions. Except in the Southern region, the model's forecast skill for the T_x has lower than the benchmark. Similar to the explanation for average temperature, the reason the model's skill is lower than the benchmark because the temperature regime of the Southern region is largely unchanged, especially in winter, despite the influence of the cold air, but only the wind regime changes, while the T_x in this region has little change. Compared to the model, the average amplitude of benchmark forecast error is larger.

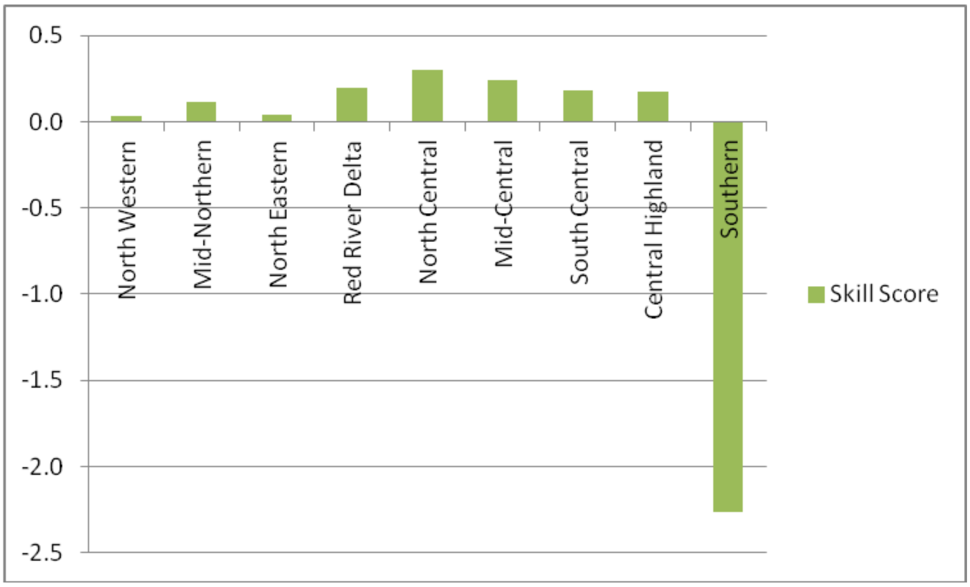

Figure 5. Skill score of IFS model for T_x at 09 regions in Viet Nam.

4. Conclusion

Through the verification of forecast error indices and forecast skill according to WMO guidance and regulations of legal documents on assessing the quality of meteorological forecasting within allowable error range about ± 2^oC for temperature, in this study we evaluate the correct and the skill of the IFS model from 2019-2022 for the minimum temperature, average temperature, maximum temperature within the 24-hours forecast period at 09 regions in Viet Nam, the results show:

- Regarding forecast bias: The IFS model tends to forecast the minimum temperature to be higher than the actual one, while the average temperature and maximum temperature to be lower than the actual one at the most regions. The average amplitude of forecast error is highest in the Red River Delta region.

- Regarding the correct: The IFS model forecast the minimum temperature and average temperature with higher correct than the maximum temperature. The correct in the southern region is higher than in the northern region, with the highest correct in the South-Central region, and the lowest correct in the Red River Delta region.

- Regarding the skill: The IFS model has better forecast skill for the minimum temperature and average temperature at the most regions. Except in the Southern region, the model's forecast skill is lower than the benchmark forecast (for average temperature and maximum temperature).

Author contribution statement: Conceived and designed the experiments; Analyzed and interpreted the data; manuscript editing: H.T.T.L; Analysis tools or data; performed the experiments: H.T.N; Wrote the draft manuscript: H.T.T.T; Contributed reagents, materials: T.T.N.

Acknowledgements: This study is supported by the funding of the project titled "Research the scientific basis to develop national technical regulations on assessment of hydro-meteorological forecasting and warning in accordance with regulations of the World Meteorological Organization (WMO) " grant number: TNMT.2023.02.34.

Competing interest statement: The authors declare no conflict of interest.

References

1. WMO. Guidelines on performance assessment of public weather services - No. 1023, 2000.
2. Murphy, A.H. What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting* **1993**, *8*, 281–293.
3. Murphy, A.H. Forecast verification. Economic value of weather and climate forecasts. Katz, R.W., Murphy, A.H. (Eds). Cambridge Univ. Press, chapter 7, 1997, pp. 19–74.
4. Thornes, J.E.; Stephenson, D.B. How to judge the quality and value of weather forecast products. *Meteorol. Appl.* **2001**, *8*, 307–314.
5. Wilks, D.S. A skill score based on economic value for probability forecasts. *Meteorol. Appl.* **2001**, *8*, 209–219.
6. Seaman, R.; Mason, I.; Woodcock, F. Confidence intervals for some performance measures of yes-no forecasts. *Aust. Met. Mag.* **1996**, *45*, 49–53.
7. Hamill, T.M. Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting* **1999**, *14*, 155–167.
8. Kane, T.L.; Brown, B.G. Confidence intervals for some verification measures - a survey of several methods. Proceeding of the 15th Conference on Probability and Statistics in the Atmospheric Sciences, Amer. Met. Soc., 8-11 May 2000, Asheville, North Carolina, 2000.

9. Hamill, T.M.; Juras, J. Measuring forecast skill: is it real skill or is it the varying climatology? *Q. J. Royal Met. Soc.* **2006**, *132*, 2905–2923.
10. Stanski, H.R.; Wilson, L.J.; Burrows, W.R. Survey of common verification methods in meteorology. World Weather Watch Tech. Rept. No. 8, WMO/TD No. 358, WMO, Geneva, 1989, pp. 114.
11. Katz, R.W.; Murphy, A.H. (Eds): Economic value of weather and climate forecasts. Cambridge University Press, Cambridge, 1997.
12. Jolliffe, I.T.; Stephenson, D.B. Forecast Verification: A practitioner's guide in atmospheric science. 2nd Edition. Wiley and Sons Ltd, 2012, pp. 274.
13. Murphy, A.H.; Katz, R.W. Probability, statistics, and decision making in the atmospheric sciences. Westview Press, Boulder, CO, 1985.
14. Murphy, A.H.; Winkler, R.L. A general framework for forecast verification. *Mon. Wea. Rev*. **1987**, *115*, 1330–1338.
15. Nurmi, P. Recommendations on the verification of local weather forecasts (at ECWMF member states). ECMWF Operations Department, 2003.
16. Hoa, V.V. Comparative study skills rain forecast the middle part and central highland of several global models. *VN J. Hydrometeorol.* **2016**, *667*, 1–8.
17. Ba, T.D.; Hoa, V.V.; Tri, D.Q. A verification of short-term rainfall forecast by using ifs model of ECMWF on the northern central region. *VN J. Hydrometeorol.* **2019**, *697*, 33–43.
18. Nga, N.T.; Thanh, C.; Hung, M.K.; Tien, D.D. Verification of quantitative rainfall forecast from IFS and WRF model for the northern region of Viet Nam. *VN J. Hydrometeorol.* **2021**, *730*, 79–92.
19. Minh, L.T.; Lam, H.P. Automatically correction for forecasts city temperature from the ifs model output. *VN J. Hydrometeorol.* **2018**, *693*, 41–47.
20. MoNRE. Circular No. 41/2017/TT-BTNMT dated October 23rd, 2017 of the Minister of Natural Resources and Environment promulgating technical regulations on assessing the quality of meteorological forecasting, 2017.