Can simple MOS bring improvement into ALADIN T_{2m} forecast?

Ivan Bašták Ďurán — Commenius University
Ján Mašek — Slovak HydroMeteorological Institute

Why do we need model output statistics (MOS)?

- outputs from NWP models are not perfect, but are subject to **errors**
- these errors can be **reduced** by:
  1. improving the numerical model (preferred way)
  2. statistical adaptation of model outputs against observations
- first approach removes the **source** of errors, but it is slower, expensive and requires joint effort of big teams
- second approach views model as **black box**, it can be implemented quickly and cheaply, but the black box should not change
- with second approach we can hope to eliminate **systematic part** of model errors

What is MOS?

- MOS = multilinear regression:

$$Y = \underbrace{\sum_{i=1}^{m} b_i X_i}_{\hat{Y}} + \varepsilon$$

Y	– predictant (observed quantity)
$\hat{Y}$	– MOS estimate of Y
$b_1, \ldots, b_m$	– regression coefficients
$X_1, \ldots, X_m$	– predictors (quantities forecasted by model, observations available at analysis time, . . .)
ε	– error of MOS estimate

- regression coefficients are determined by least squares method, i.e. by minimization of mean square error $\overline{(\hat{Y} - Y)^2}$ on **training** data set

- MOS skill is evaluated using independent **testing** data set

MOS limitations

- number of predictors must be **much smaller** than size of training data set (selection of too many predictors leads to overfitting)
- training period should be **sufficiently long** (in ideal case 5 years or more) in order to correctly sample different weather situations
- time series of model outputs should be **homogeneous** (numerical model should not change during period of MOS training and usage)

Questions to be answered

1. Can simple MOS improve ALADIN T_{2m} forecast despite frequent model changes?

2. What would be optimal design of the MOS system?

3. Can more sophisticated MOS bring substantial improvement compared to simple MOS?

Used data

- studied period: 2000–2004 (5 years)

- observations:

  SYNOP T_{2m} observations from 9 Slovak stations

- forecasts:

  ALADIN pseudoTEMPs (forecasted vertical profiles of pressure, temperature, humidity and wind)

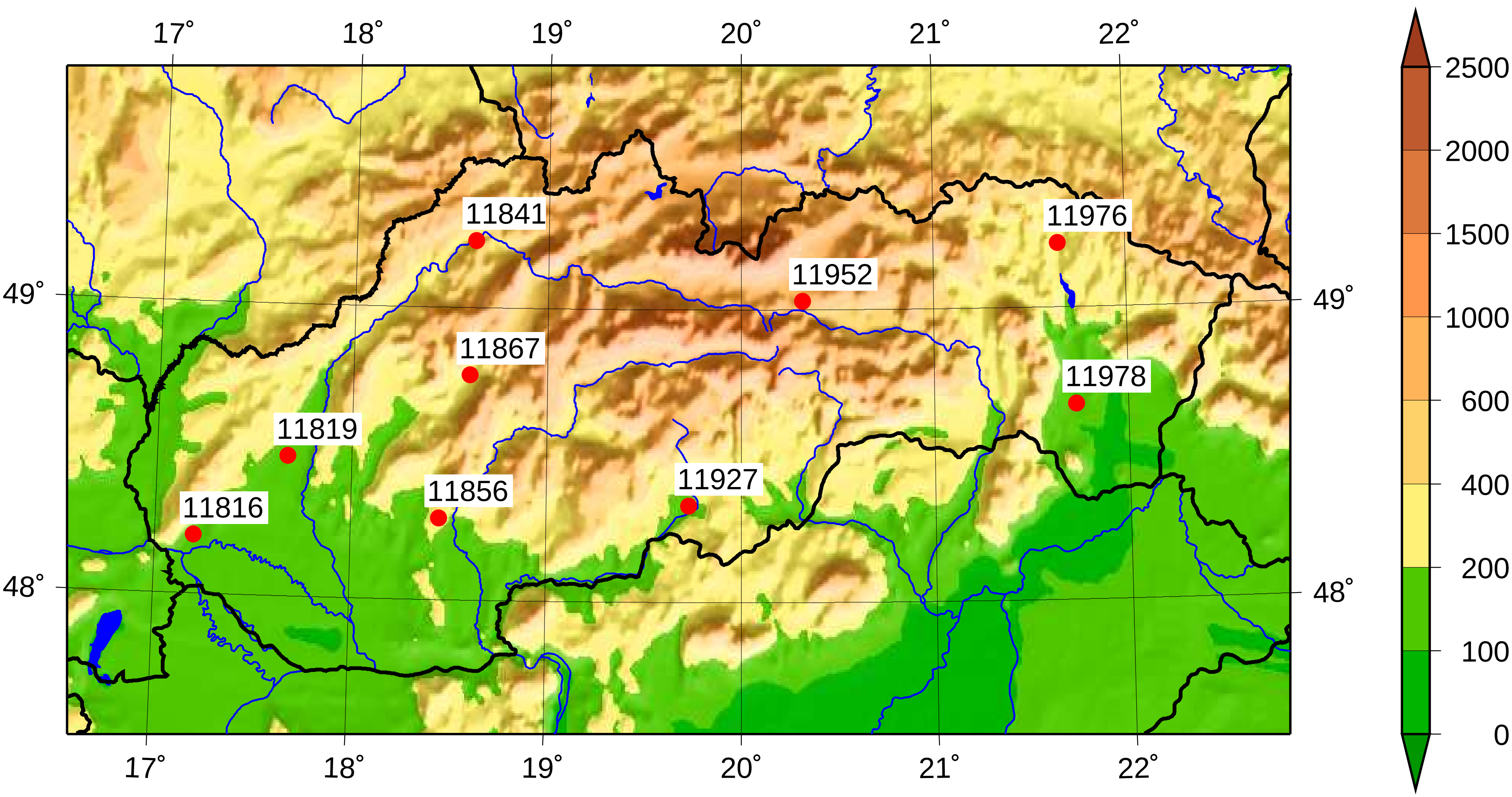
  – used operational models:

Jan 2000 – Dec 2002	ALADIN/LACE	Prague
Jan 2003 – Jun 2004	ALADIN/LACE	Vienna
Jul 2004 – Dec 2004	ALADIN/SHMU	Bratislava

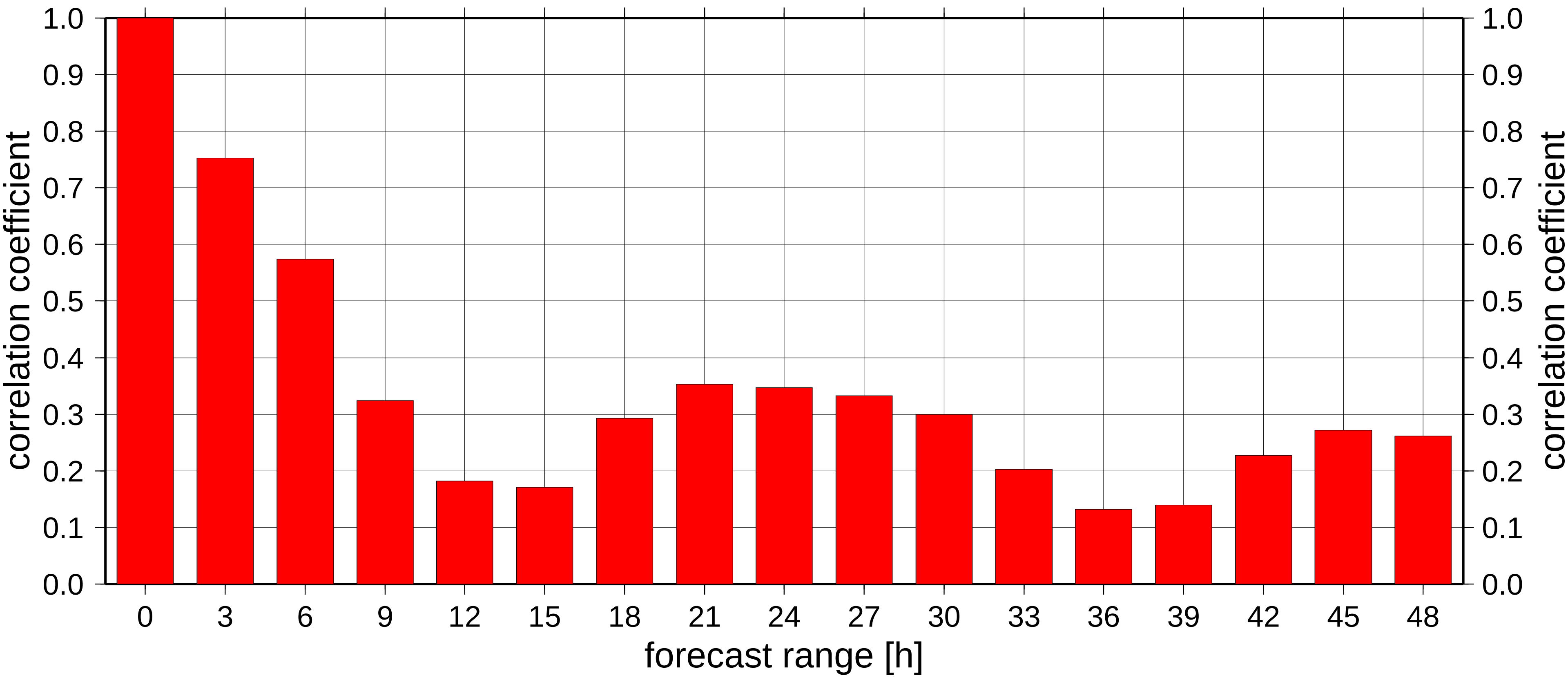
  – restriction to 00 UTC integration

  – concentration on +36 h forecast

Selected stations:

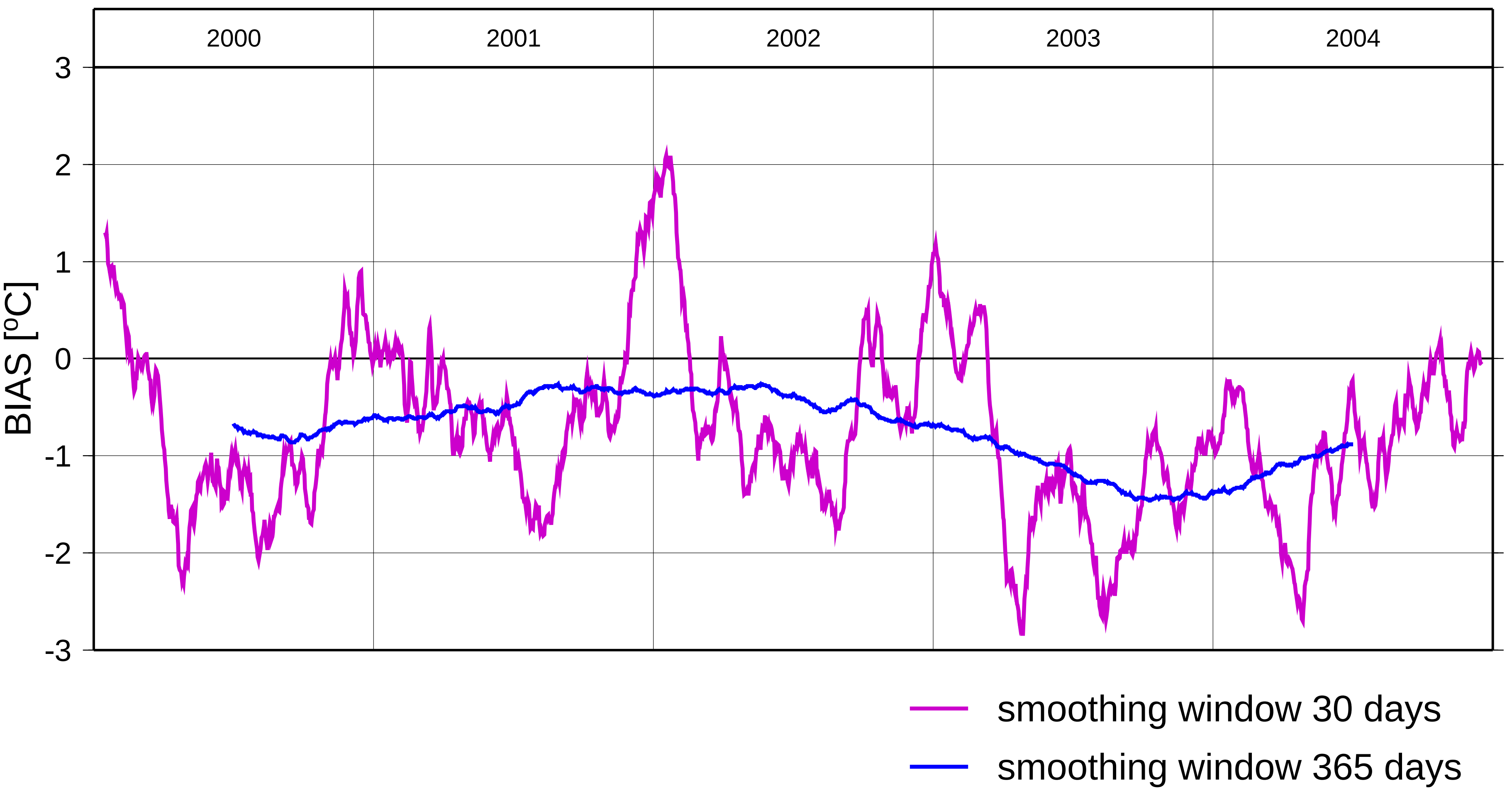


Autocorrelation function of model T_{2m} error (forecast against analysis):



(period 2000–2004, 00 UTC integration, average over all stations)

Evolution of T_{2m} BIAS:



(00 UTC integration, +36 h forecast, all stations)

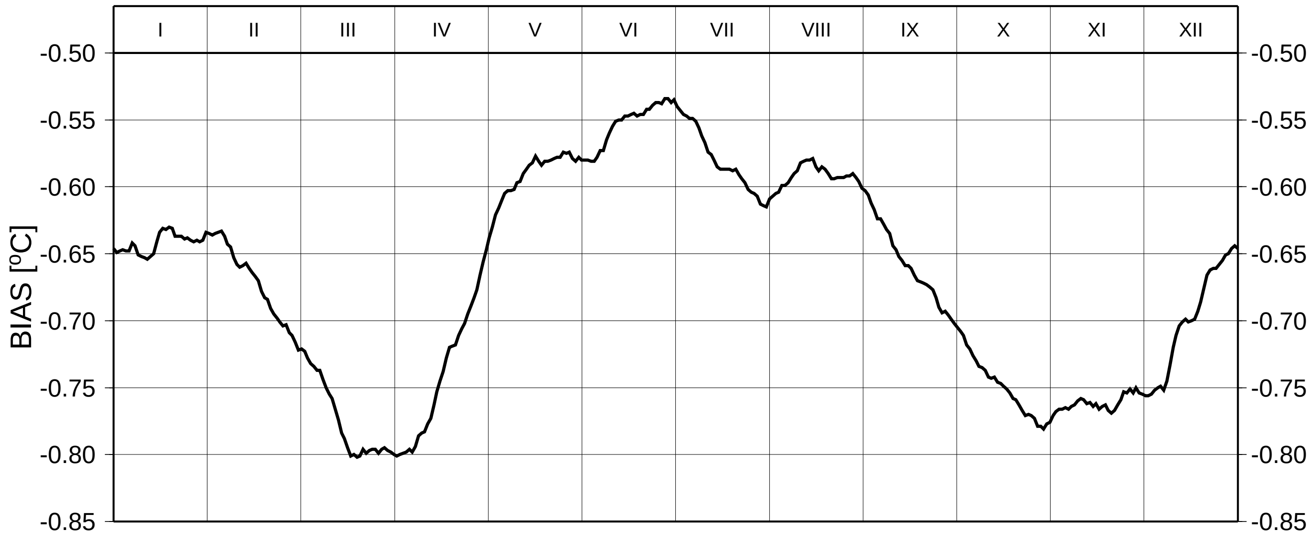
Design of simple MOS

- separate regression model for each station
- predictant: error of model T_{2m} forecast $(T_F^+ - T_O^+)$
- predictors: 1, error of model T_{2m} analysis $(T_F^0 - T_O^0)$, $\cos\theta$, $\sin\theta$, $\cos 2\theta$, $\sin 2\theta$; where θ is time of year (goes from 0 to 2π)
- time predictors $\cos\theta$, $\sin\theta$, $\cos 2\theta$ and $\sin 2\theta$ are included in order to describe annual course of model BIAS
- alternative way is to cluster data into several groups according to part of year and develop separate MOS for each group:

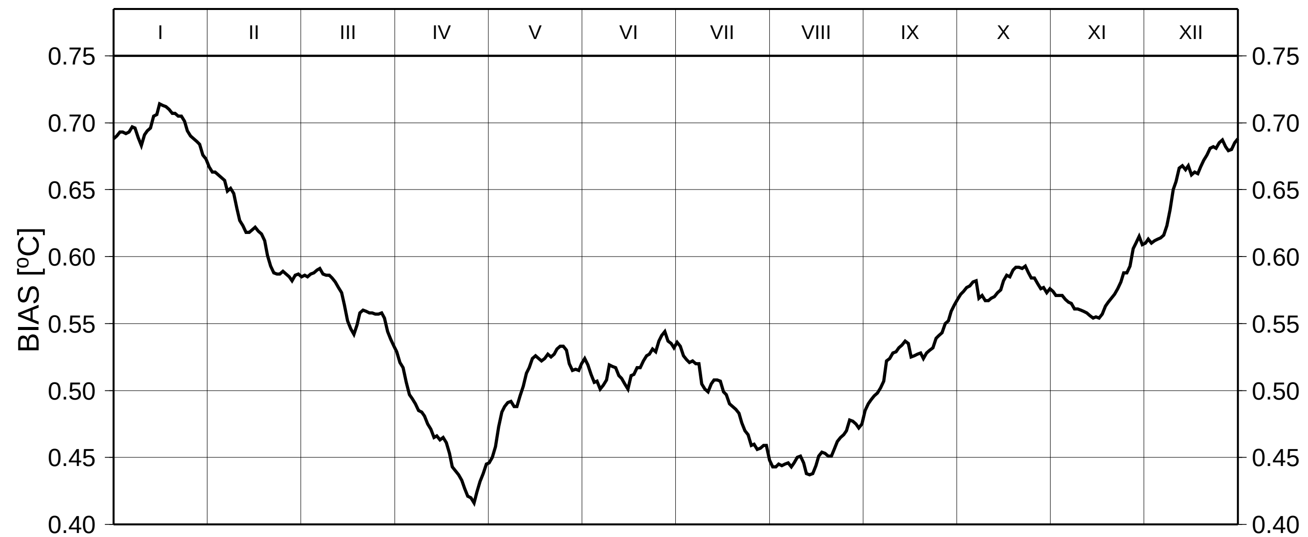
training												testing			
2000				2001				2002				2003			
	●				●				●				●		

Annual course of T_{2m} BIAS:

+24 h forecast:



+36 h forecast:



(period 2000–2004, 00 UTC integration, all stations)

Tested configurations

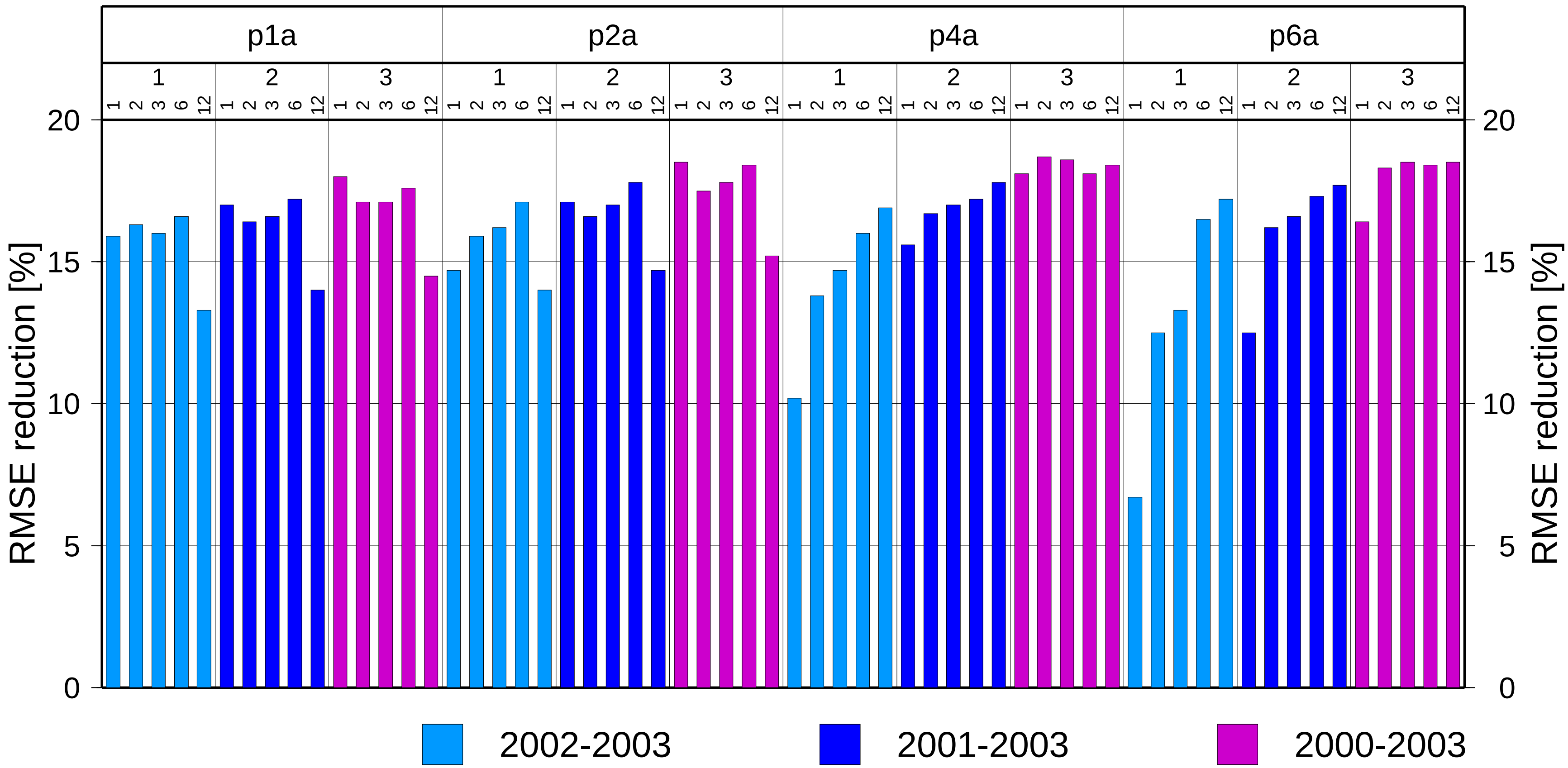
- predictor selections:
  - p1a . . . 1 (simple BIAS correction)
  - p2a . . . 1, $T_F^0 - T_O^0$
  - p4a . . . 1, $T_F^0 - T_O^0$, $\cos\theta$, $\sin\theta$
  - p6a . . . 1, $T_F^0 - T_O^0$, $\cos\theta$, $\sin\theta$, $\cos 2\theta$, $\sin 2\theta$

- time window for data clustering: 1, 2, 3, 6 and 12 months (12 months means no clustering)

- training period: 1, 2 and 3 years

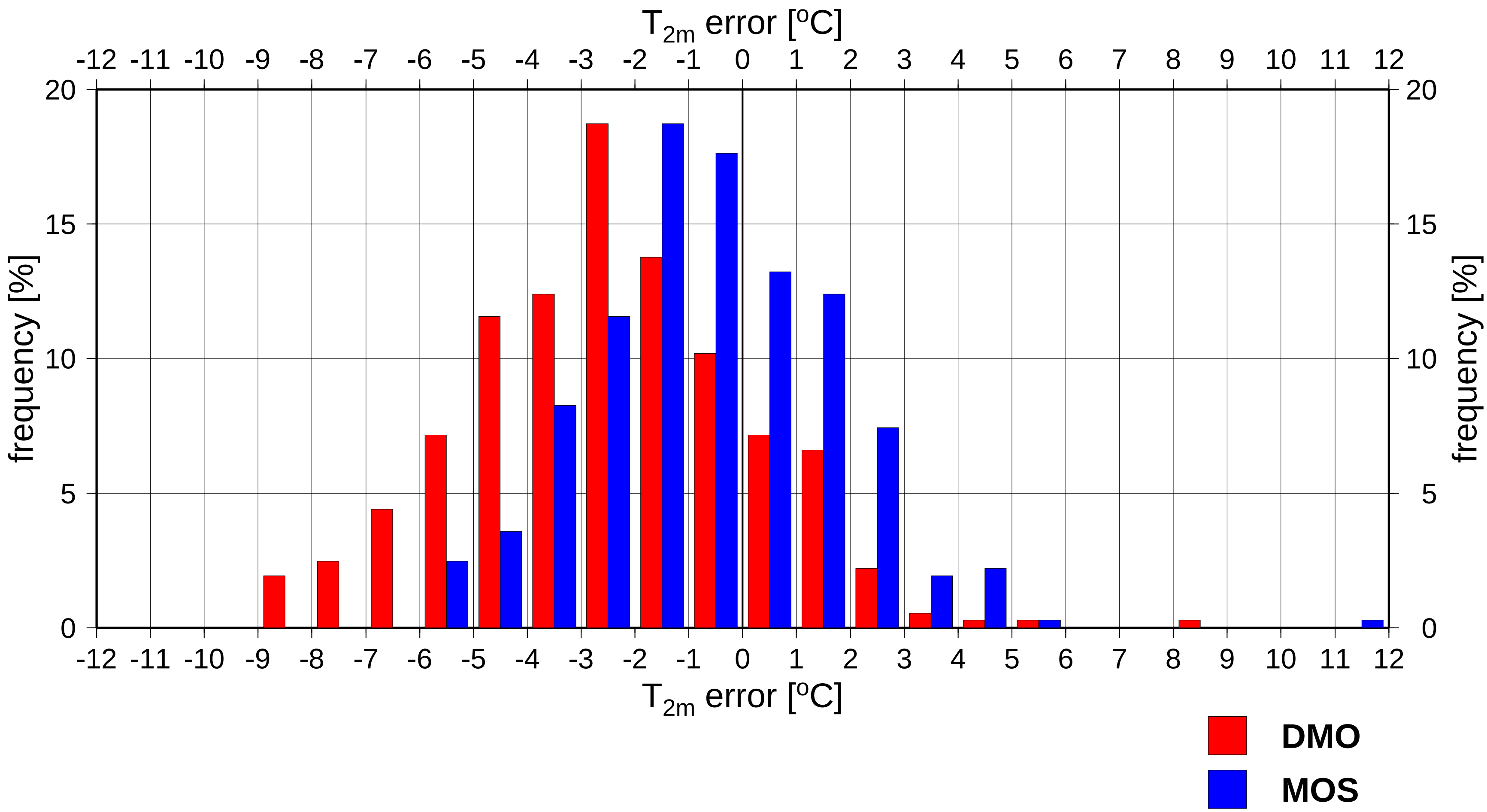
- testing period: 1 year

T_{2m} RMSE reduction, testing year 2003:



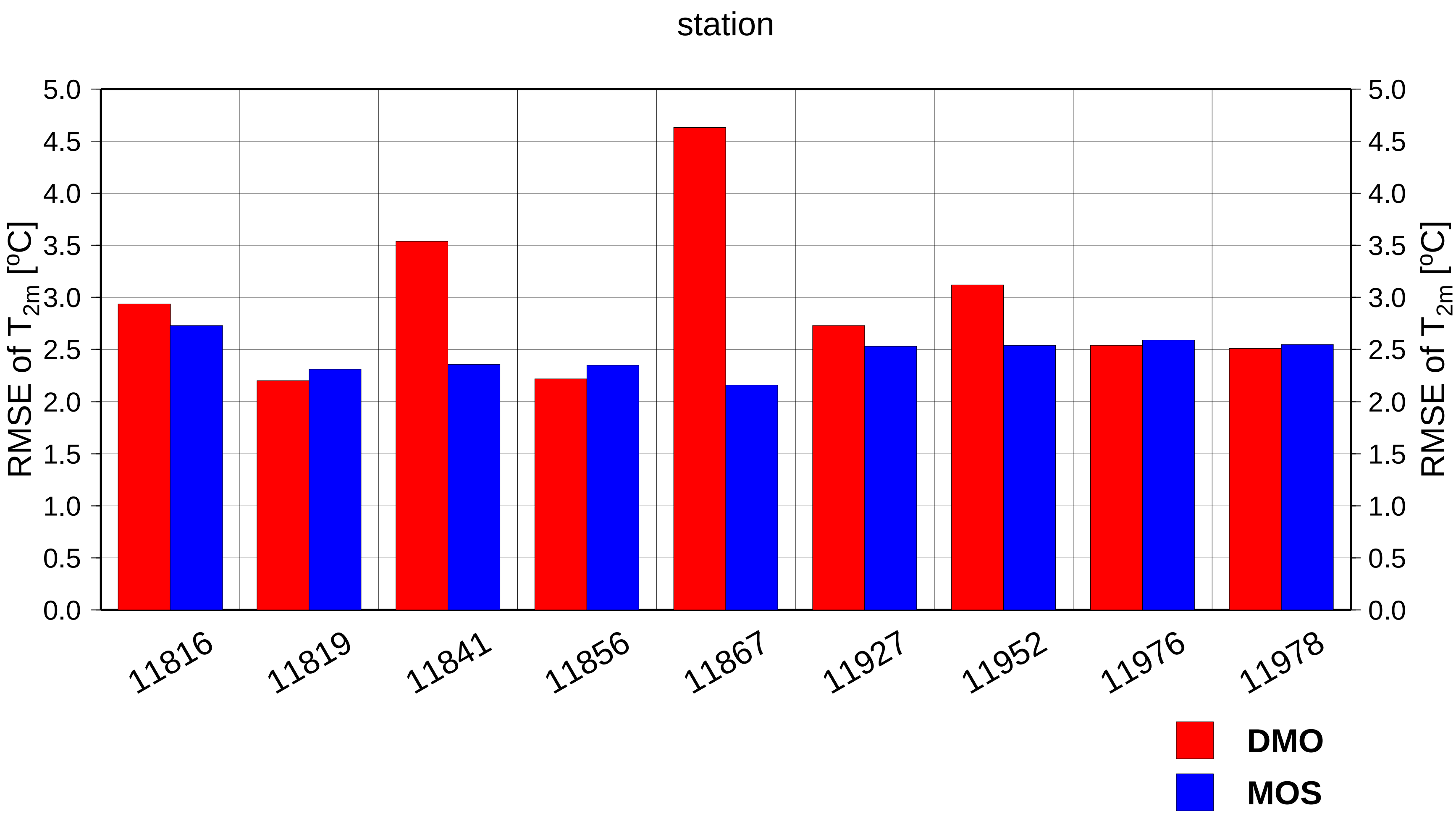
(00 UTC integration, +36 h forecast, all stations)

T_{2m} error distribution for station 11841 Žilina, testing year 2003:



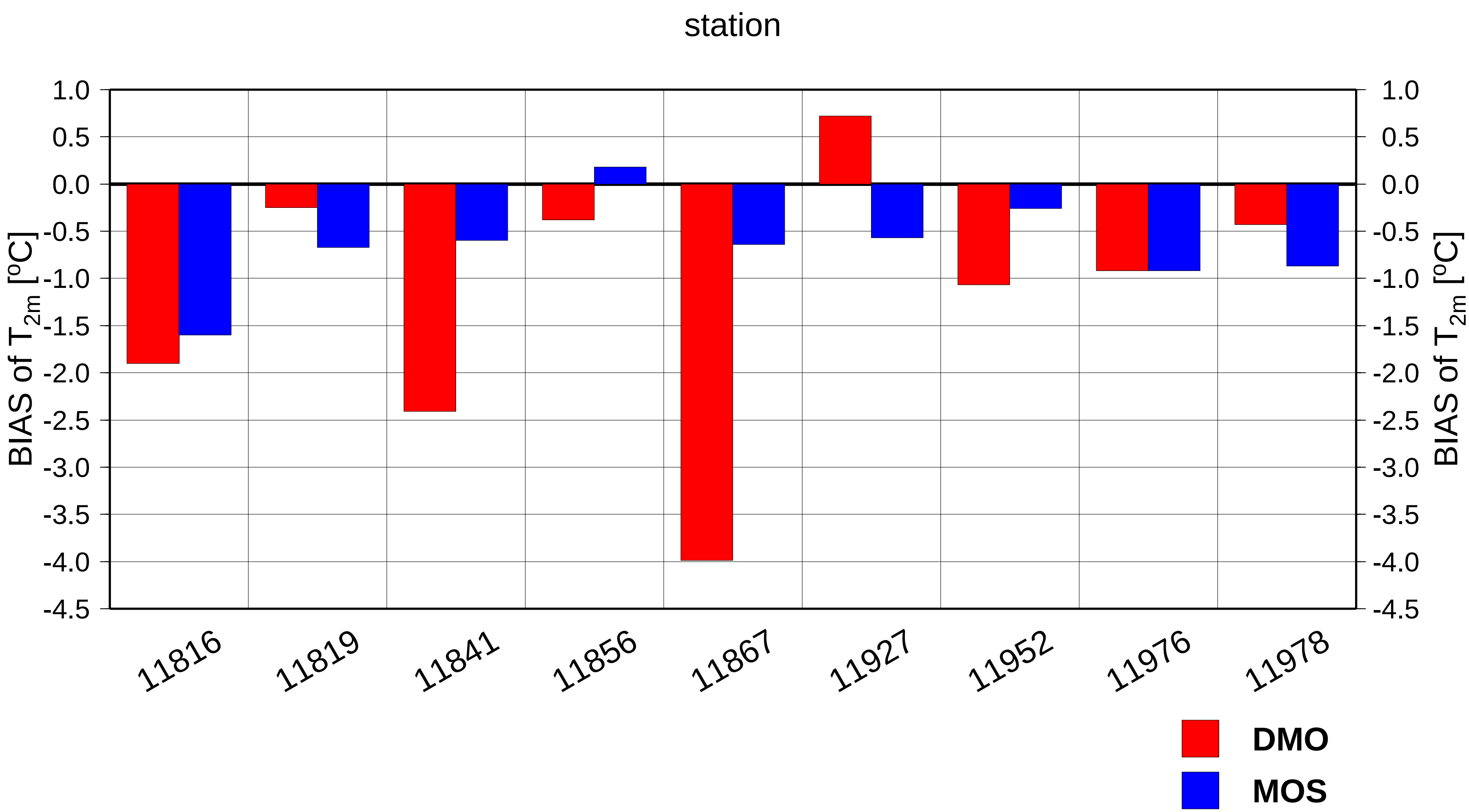
(predictor selection p6a, training period 2000-2002, time window 12 months, 00 UTC integration, +36 h forecast)

T_{2m} RMSE for individual stations, testing year 2003:



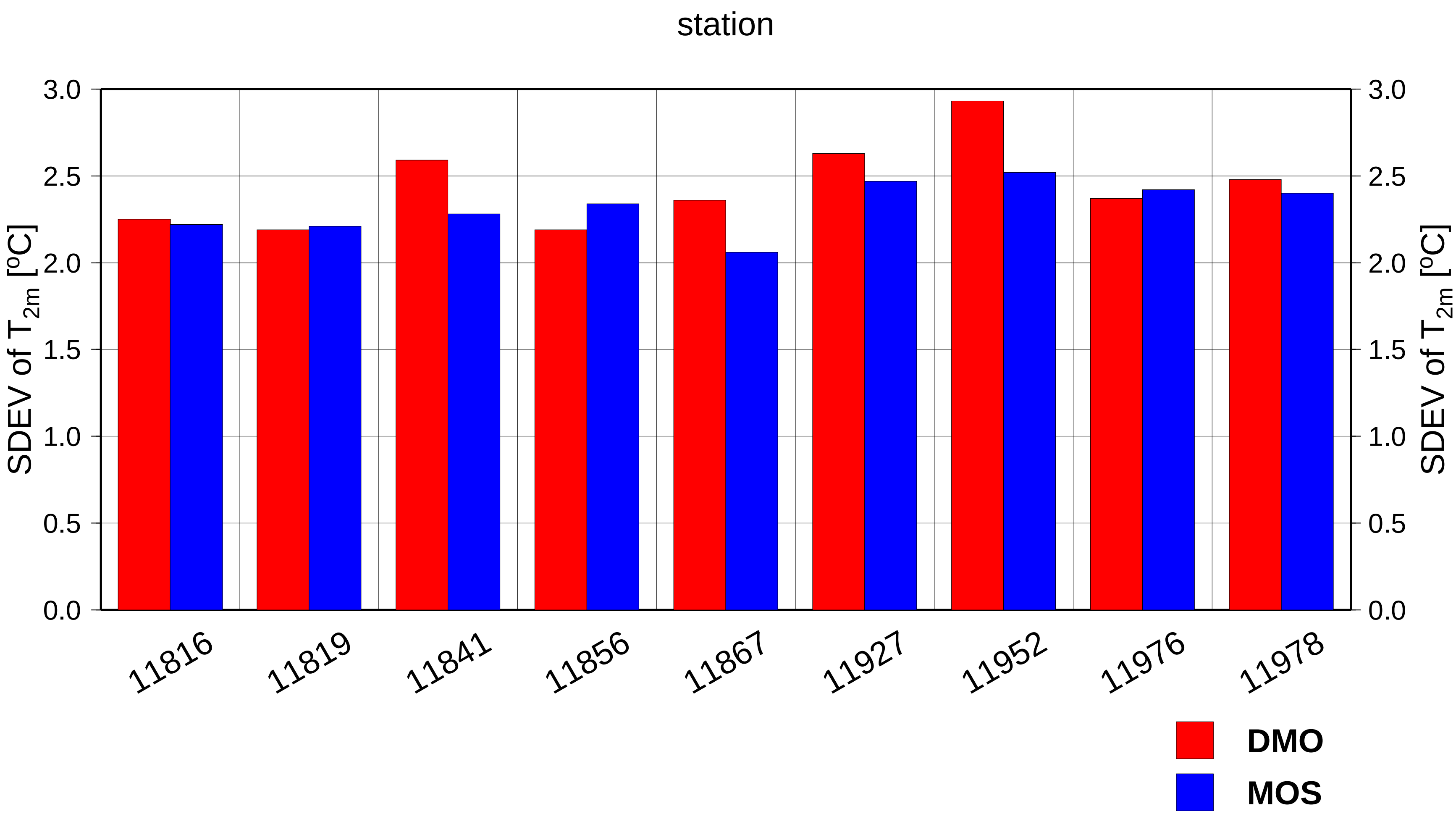
(predictor selection p6a, training period 2000-2002, time window 12 months, 00 UTC integration, +36 h forecast)

T_{2m} BIAS for individual stations, testing year 2003:



(predictor selection p6a, training period 2000-2002, time window 12 months, 00 UTC integration, +36 h forecast)

T_{2m} SDEV for individual stations, testing year 2003:

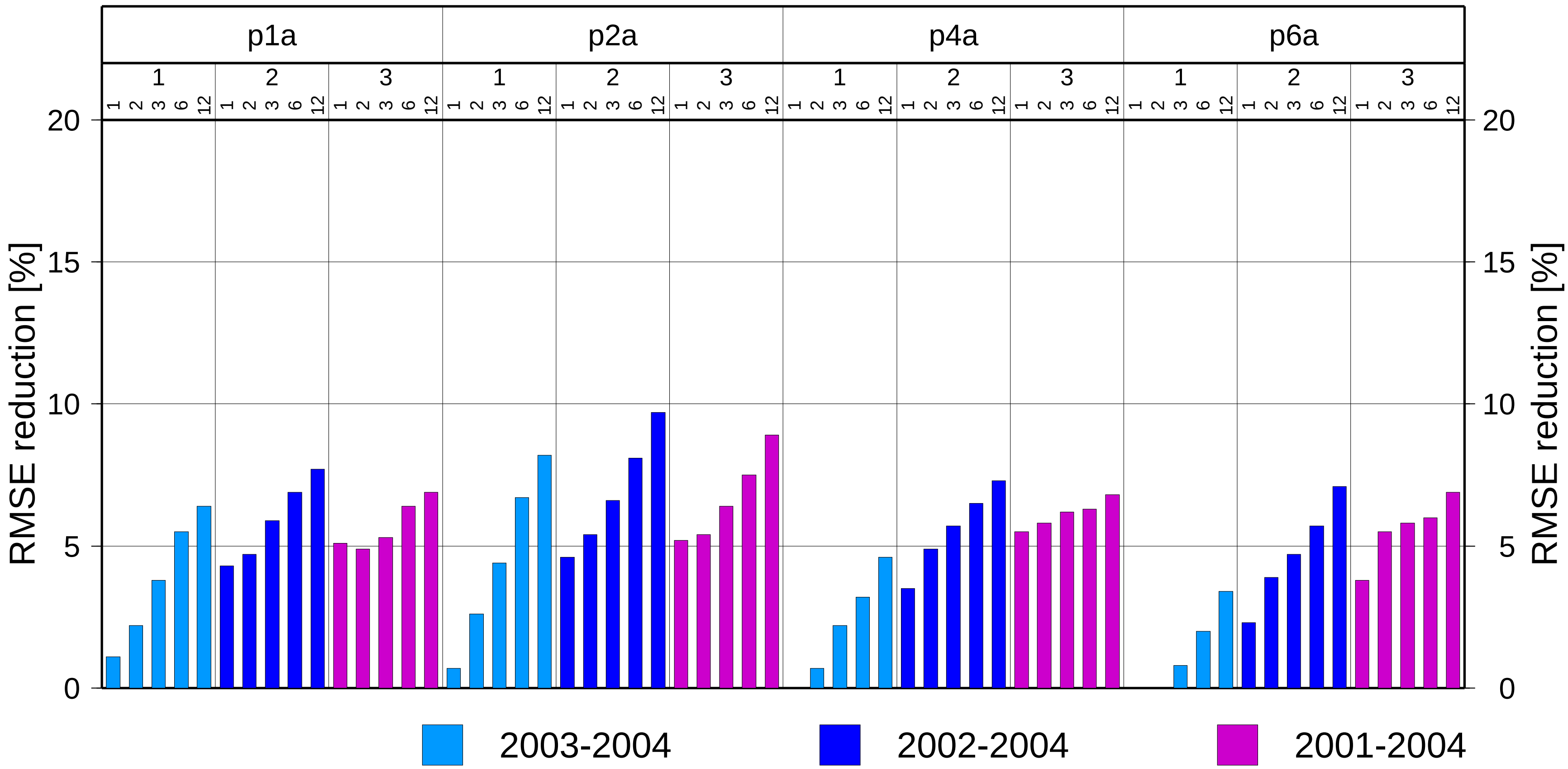


(predictor selection p6a, training period 2000-2002, time window 12 months, 00 UTC integration, +36 h forecast)

Results from simple MOS I

- the longer training period, the better MOS results
- including of analysis error among predictors improves MOS skill compared to simple BIAS correction
- data clustering **or** use of time predictors improves MOS performance
- combination of data clustering with use of time predictors leads to overfitting especially for short time windows and short training periods
- RMSE reduction is achieved by correcting yearly BIAS
- for best configurations overall RMSE reduction reaches 18%
- the most attractive candidate seems to be: p6a, training period 3 years, time window 12 months

T_{2m} RMSE reduction, testing year 2004:



(00 UTC integration, +36 h forecast)

Results from simple MOS II

- results from testing year 2004 are bad surprise
- previously selected optimal configuration reaches overall RMSE reduction only 7%
- longer training period does not imply better MOS performance
- data clustering or use of time predictors deteriorate MOS results
- best configuration is now: p2a, training period 2 years, time window 12 months; with overall RMSE reduction 9%

Cause of simple MOS failure

- during period 2000–2003 there were many changes in operational model ALADIN:

  al11 → al12op3 → al15 → al25t2
  different physical parametrisations and their tunings
  (CYCORA, CYCORA_bis, CYCORA_ter+++)
  dynamical adaptation, blending
  6 h → 3 h coupling frequency, 31 → 37 vertical levels

- however, there was no change in horizontal geometry

- in 2004, model resolution changed from 12.2 km to 9.0 km

- it seems that related change of model orography is a critical factor for behaviour of T_{2m} error

- this is not surprising, since there is significant altitude dependence of T_{2m}

Design of more sophisticated MOS

- possibility of regionalized regression model (i.e. common regression equation for all stations with geographical predictors added)

- predictant: observed T_{2m} at forecast time

- potential predictors:

  1, observed T_{2m} at analysis time

  forecasted quantities p, T, r, u, v and $\sqrt{u^2 + v^2}$ at heights 2, 20, 200 and 2000 m above model surface

  time predictors: $\cos\theta$, $\sin\theta$, $\cos 2\theta$, $\sin 2\theta$

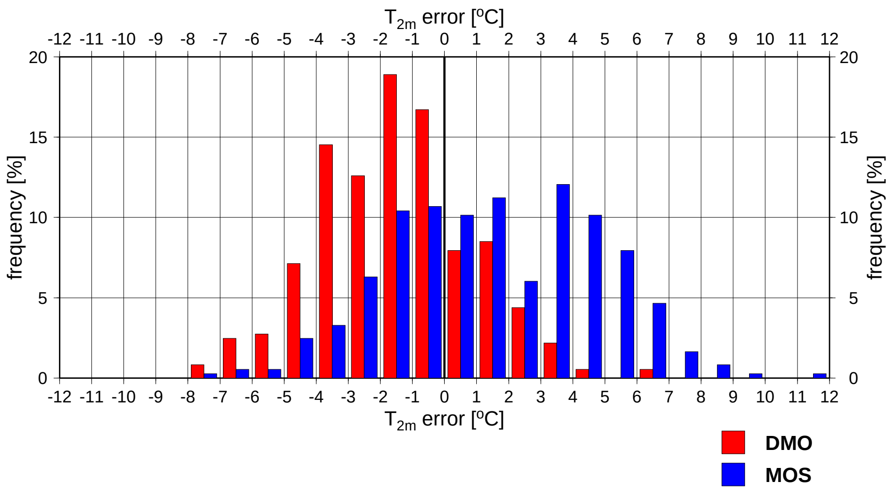
  geographical predictors: λ, φ, $h_{model} - h$

- possibility of data clustering (less attractive alternative to the use of time predictors, since it reduces size of training data set)

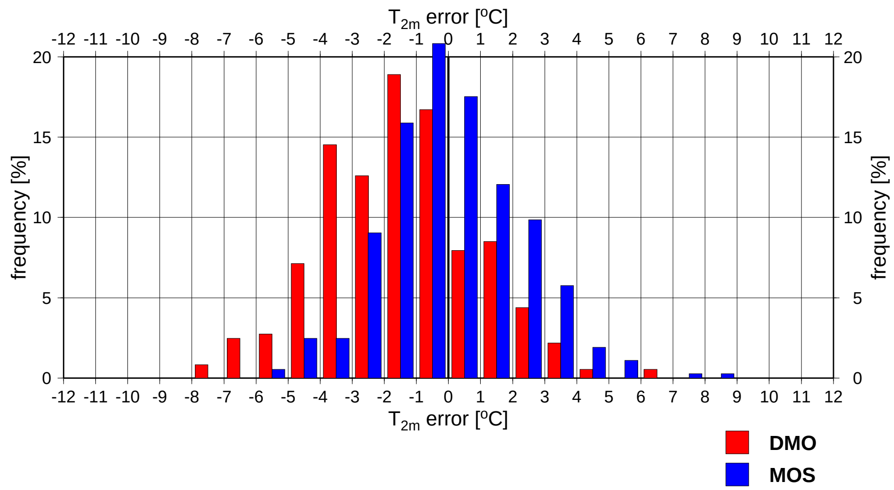
- selection of final predictors by forward screening

T_{2m} error distribution for station 11841 Žilina, testing year 2004:

individual MOS equations:



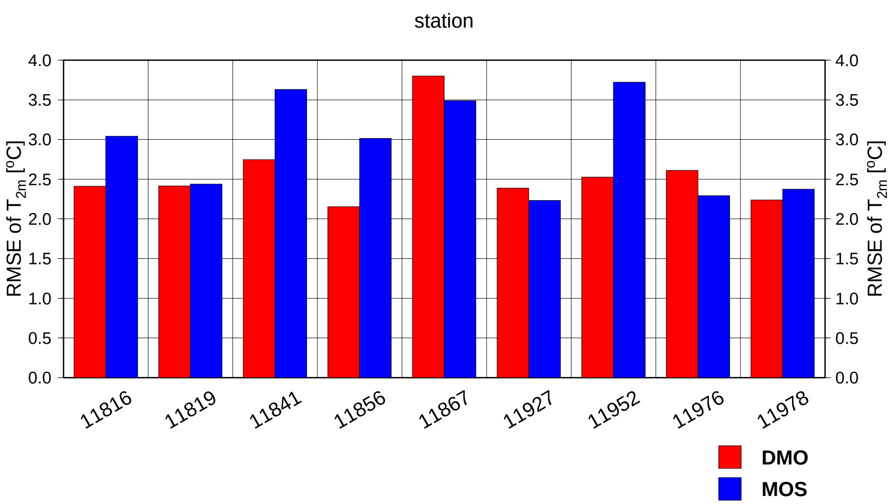
regionalized MOS equation:



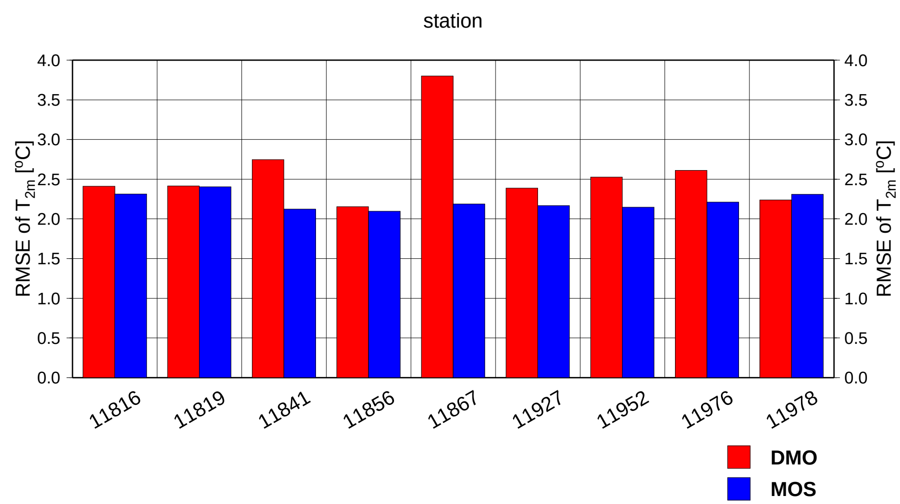
(training period 2001–2003, time window 3 months, no time predictors, 00 UTC integration, +36 h forecast)

T_{2m} RMSE for individual stations, testing year 2004:

individual MOS equations:



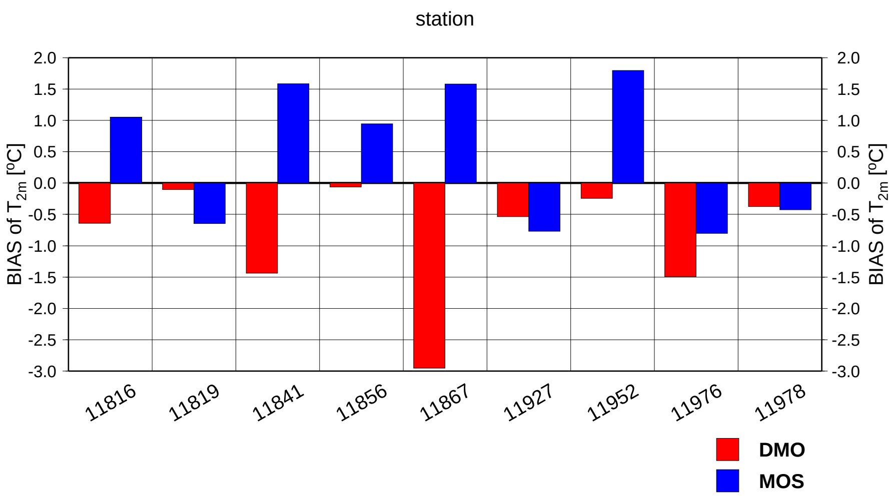
regionalized MOS equation:



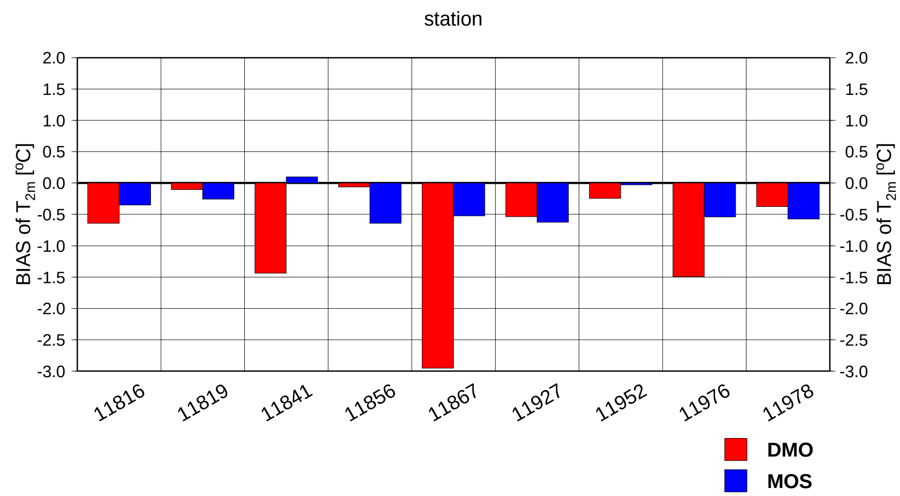
(training period 2001–2003, time window 3 months, no time predictors, 00 UTC integration, +36 h forecast)

T_{2m} BIAS for individual stations, testing year 2004:

individual MOS equations:



regionalized MOS equation:



(training period 2001–2003, time window 3 months, no time predictors, 00 UTC integration, +36 h forecast)

Results from more sophisticated MOS

- when model orography does not change, individual MOS equations give better results than regionalized MOS (not shown)
- individual MOS equations are not usable when model orography changes (for shown configuration overall RMSE **increased** by 13%, results are even worse than for simple MOS)
- however, regionalized MOS equation is usable despite the change of model orography (for shown configuration overall RMSE **decreased** by 15%)
- regionalized MOS equation reduces RMSE mainly by correcting yearly BIAS
- there is slight reduction of yearly SDEV, probably thanks to fitting the annual course of model BIAS (not shown)

Conclusions

- at current horizontal resolutions ($\approx 10\,\text{km}$), MOS still **can improve** T_{2m} forecast
- strongest limitation does not come from modifications of physical parametrizations, but from changes of **model orography**
- in order to cope with this problem, **regionalized approach** must be used, including $h_{model} - h$ among predictors
- care must be taken to **selection of predictors**, since overfitting can occur even if only few predictors used
- preferable way how to describe annual course of model BIAS is to use **time predictors**
- except from effect of regionalization, more sophisticated MOS **does not** bring much improvement compared to simple MOS