Multi-model time-lagged ensemble (ICMENS) – results of the LOBSTER project

Marta Kopeć[1], Magdalena Gruziel-Słomka[1], Anna Jagodnicka[1]

[1] Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw

Atmospheric Physics Seminar, 08/03/2024

Outline

1. LOBSTER project;
2. Multi-model ensemble forecasts
3. UM/GFS example verification;
4. Model fitting
5. ICMENS example verification;
6. ICMENS automatic setup;
7. Future plans, summary and conclusions;

LOBSTER project - brief description:

LOBSTER is a system supporting energy management at **Local Balancing Areas** with dispersed renewable generation, energy storage, and e-mobility infrastructure.

It was conducted at ICM UW (with cooperation with Globema Company) between 2021 and 2023 and it was was granted funding by NCBiR (The National Centre for Research and Development; POIR.01.01.01-00-0507/20).

globema

European Funds Smart Growth

Republic of Poland

NCBR National Centre for Research and Development

European Union European Regional Development Fund

Main goals of the project:

- Create controllable Local Balancing Areas;
- Maximize RES (Renewable Energy Sources) production;
- Better energy production and weather forecasts.

Why do we want multi-model time-lagged ensembles:

- to improve forecasts;
- to elongate them (consistent forecast for many days with better performance than single, global models);
- to reduce of calculation costs;

ICMENS: brief description, part 1

ICMENS is a two-model, time-lagged ensemble forecast. Models that are currently used:

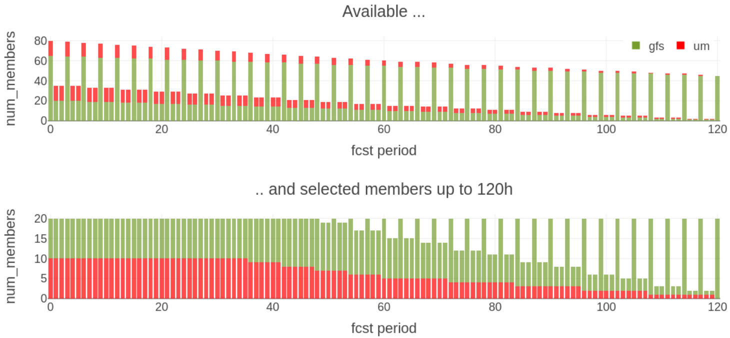
- regional, 4km UM vn10.1 (developed by MetOffice in UK, run at ICM UW)
- global, 20km GFS (developed in National Centers for Environmental Prediction, USA; results available for download)

Both models produce forecasts every 6h:

- UM forecasts: hourly resolution; 120h long (at 00UTC, 12UTC), or 60h long (at 06UTC, 18UTC)
- GFS forecasts: hourly resolution up to 120h, followed by 3-hourly up to 384h (16 days)

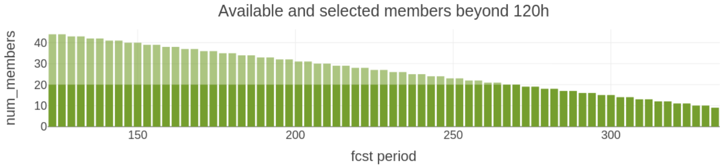
ICMENS: brief description, part 2

ICMENS produces 14-day forecast every 24h (at 00 UTC) with hourly resolution up to 120h, then 3-hourly afterwards. Only a maximum of 20 members are used for a single forecast.



ICMENS: brief description, part 2

ICMENS produces 14-day forecast every 24h (at 00 UTC) with hourly resolution up to 120h, then 3-hourly afterwards. Only a maximum of 20 members are used for a single forecast.



ICMENS: basic concept

The basic idea is to reduce the additive bias of each ensemble member. Thus the final forecast for a period t_i is either (current approach) a median or (first approach) an unweighted average over all members valid at that time, each corrected by mean error:

$$X(t_i) = \frac{1}{M} \sum_{m=1}^{m=M} X_m^{corr}(t_i) \tag{1}$$

(M is the number of available members, and $X_m^{corr}(t_i) = X_m(t_i) - e_m(t_i)$, e - mean error)

The correction is obtained from Random Forrest Regression (RFR) model run on the results of yearly verification of both models (UM, GFS).

UNIWERSYTET WARSZAWSKI
Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego
icm.edu.pl

Verification (performed once)

1. Perform verification
2. Fit RFR model with the resulting additive bias (mean error)

Ensemble forecast (cyclic)

3 - ... Correct current forecast members with predicted mean error and combine them

(Disclaimer: project-wise development; one has to deliver working solution by deadline, hence, many decisions were made almost ad hoc; topic left for final slides)

A yearly verification (entire 2022) of selected fields from GFS and UM models against gridded ERA5[1] reanalysis data was performed. `metplus` wrapper for `GridStat` tool from MET[2] software was used.

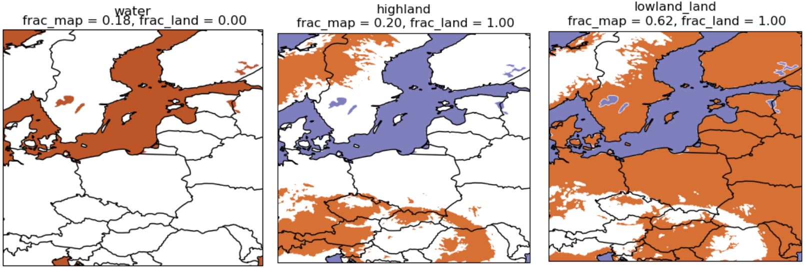
Fields verified:

- mean sea level air pressure,
- **near surface air temperature (at 1.5-2m)**,
- near surface relative humidity,
- wind speed at 10m ..,
- and 100m above ground surface.

MET
Model Evaluation Tools

[1] https://cds.climate.copernicus.eu
[2] https://dtcenter.org/community-code/metplus

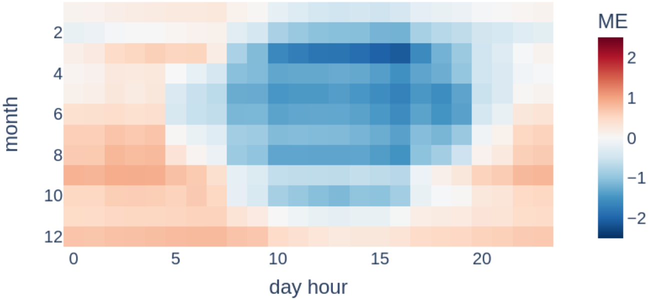
- A common 400x440 grid of 0.05° resolution (ca 5km) assumed
- Verification performed for full domain and in subdomains: **water**, NE-, NW-, SE-, SW-quarters, **Poland, not-Poland land**[1], highland (surface height $\geq$ 300m), and lowland land (the complementary to highland region, with exception of region over the water).



[1] these subdomains are used in currently running setup

UM verification: air temperature

Daily mean error (in °C) averaged over full domain, all leads, all valid times:

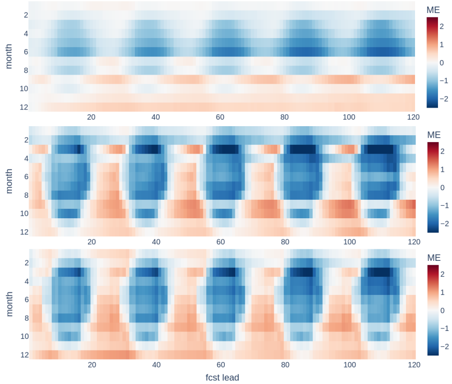


- clear daily mean error fluctuations (model too warm at night and too cold during the day)
- clear monthly "dependence" of daily mean error fluctuations

UM verification: air temperature

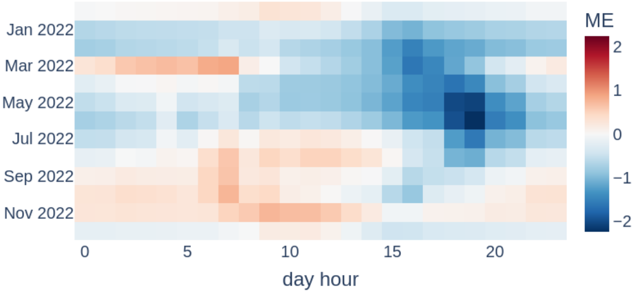
Mean error vs forecast lead hour for water, Poland and not-Poland subdomains, all valid times, only forecasts starting at 00 UTC initial time:

- daily fluctuations (with amplitude ca 2.5°) dominate; the amplitude increases with forecast lead
- smaller bias over water subdomain (close to 0 for Dec, Jan, Feb)
- largest bias .. over Poland domain



GFS verification: air temperature

Daily ME averaged over full domain, leads ≤120h, all valid times:

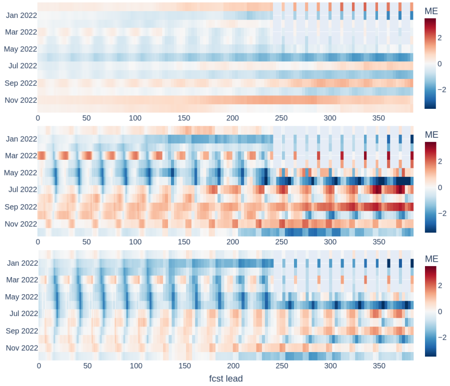


- daily bias fluctuations with minima in the afternoon
- monthly "dependence" of minima location (later in the summer)

UNIWERSYTET WARSZAWSKI
Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego
icm.edu.pl

Mean error vs forecast lead hour for water, Poland and not-Poland subdomains, all valid times, only for forecast from 00 UTC initial time:

- daily fluctuations more clear; the bias trend increases with forecast lead
- smallest bias over water
- largest bias again over Poland



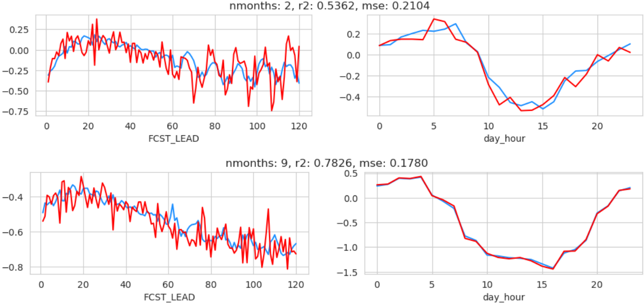
Model fitting

1. Uses mean error calculated with `metplus GridStat`:
   a) ERA5 data used as the reference;
   b) several subregions (PoNPLoW);
2. Obtained statistics are used as an input data for **Random Forest Regression** (RFR) model[2].
   The model is fitted separately for each subregion, for each variable. Features used for prediction are: field value, lead hour, day hour, month, and forecast init hour.
3. The final forecast is a median of ensemble members corrected with usage of predicted (by RFR) ME.

[2] scikit-learn Python package was used

Random forest regression - a forest of decision trees; instead of a single tree, several trees are built typically using selection of predicting features.

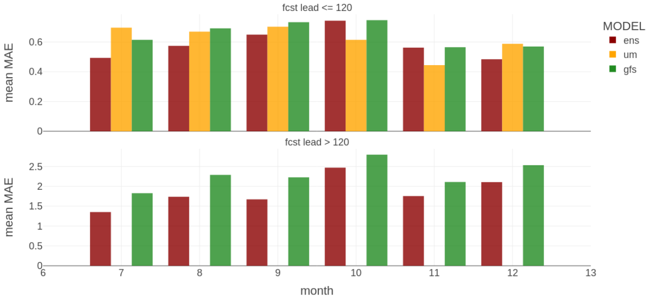
Example aggregated results of train (blue) and test sets with 2 (top) and 9 months of UM air temp ME data used.



ICMENS vs point observations (air temp)

Mean absolute error from point observations[2] relatively to ERA5
Results obtained with `PointStat` tool from `metplus` collection.

- monthly aggregates (ICMENS typically not worse..; slightly corrects errors beyond 120h):

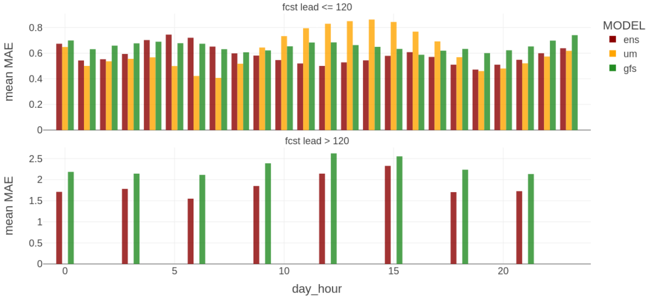


[2] https://danepubliczne.imgw.pl/

ICMENS vs point observations (air temp)

Mean absolute error from point observations[2] relatively to ERA5
Results obtained with `PointStat` tool from `metplus` collection.

- day hour aggregates (ICMENS reduces error during day hours, performs poorly for the rest of the day; slightly corrects errors beyond 120h):

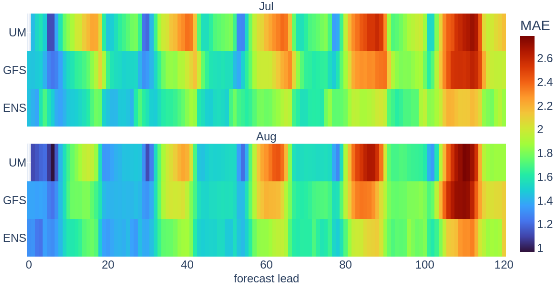


[2] https://danepubliczne.imgw.pl/

ICMENS vs point observations (air temp)

Mean absolute error - forecast lead aggregates

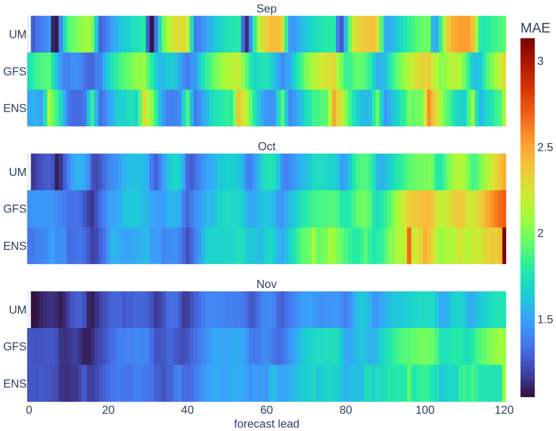
- summer months: ICMENS slightly better than UM or GFS



ICMENS vs point observations (air temp)

Mean absolute error - forecast lead aggregates

- autumn months: ICMENS not better

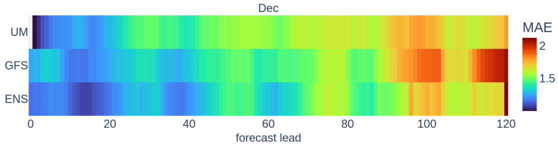


ICMENS vs point observations (air temp)

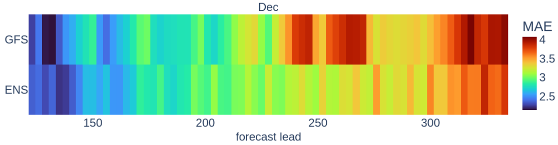
UNIWERSYTET WARSZAWSKI
Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego
icm.edu.pl

Mean absolute error - forecast lead aggregates

- winter (only Dec): ICMENS better than UM or GFS...



.. also beyond 120h:

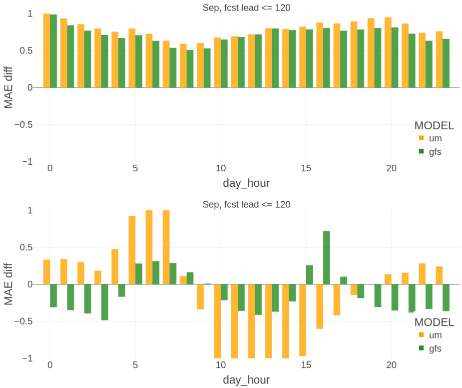


ICMENS vs grid observations (air temp)

Mean absolute error (MAE) difference between ICMENS and UM or GFS models, over two domains: **water** and **Poland**.

September, 2023 (forecast lead ≤120h)

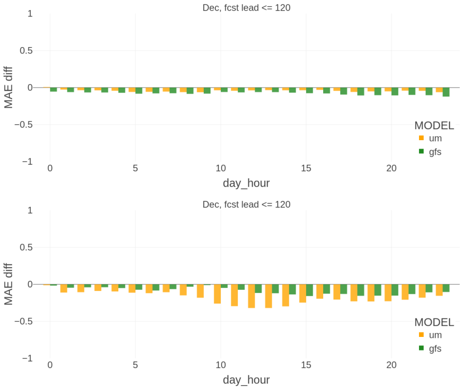
- water domain - spoiled
- Poland - ICMENS corrects mid-day hours, spoils early morning



Mean absolute error (MAE) difference between ICMENS and UM or GFS models, over two domains: **water** and **Poland**.

December, 2023 (forecast lead $\leq$120h)

- ICMENS slightly better than any of the models

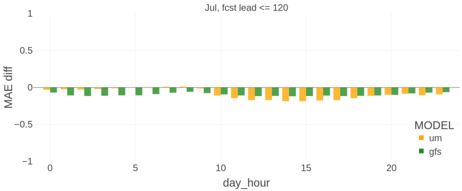


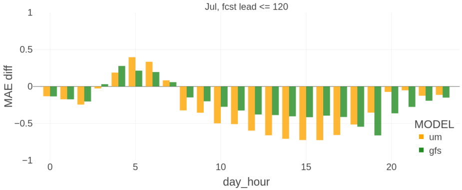
ICMENS vs grid observations (air temp)

Mean absolute error (MAE) difference between ICMENS and UM or GFS models, over two domains: **water** and **Poland**.

July, 2023 (forecast lead $\leq$120h)

- ICMENS slightly better over water
- ICMENS better over Poland except for 3-7 am

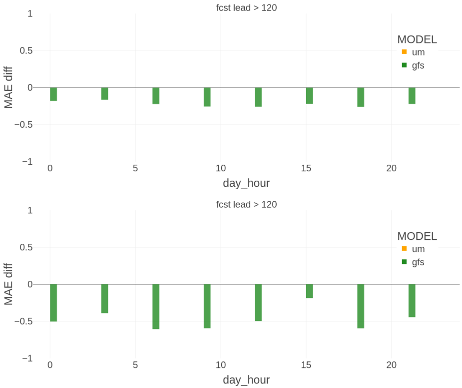




Mean absolute error (MAE) difference between ICMENS and UM or GFS models, over two domains: **water** and **Poland**.

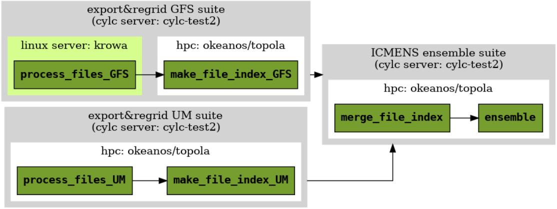
July-December, 2023 (forecast lead beyond 120h):

- ICMENS reduces the error over both domains



UNIWERSYTET WARSZAWSKI
icm Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego
icm.edu.pl

- workflow automated with `cylc`[2], vn 8.1.
- 2 suites for preprocessing current GFS and UM model results (run every 6h)
- 1 suite for the ensemble (runs at T00) triggered when the upstream preproc suites finish and the file indices are created



[2] https://cylc.github.io/

Discussion: future plans

Improvements:

- in general approach (e.g. drop the bias correction idea?)
- or keep the general approach, but change the input (e.g. use more verification subdomains classified according to daily mean error fluctuations (work in progress;)? use other models?)
- in the mean-error prediction algorithm (other method? e.g. Quantile Regression Forest, add more years of verification? recursive prediction of ME for each forecast hour based on ME from previous hours (work already in progress)? individual approach for each model, each field..? cross-correlation between fields mean error (e.g. air temperature and relative humidity)?)
- or maybe simplify? use UM members up to 120h, then GFS (e.g. corrected with QRF prediction of mean error)

Extensions:

- use more/other models: GEFS (would add more members)?, UM 1.5km (would allow for increased resolution)?
- obvious (though not that easy): generate more fields (in particular in the context of RES)
- make the ICMENS results available to broader public (e.g. meteo.pl webpage)

Benefits:

- ICMENS model improves the long-term forecast (beyond 120h),
- ICMENS reduces the mid-day bias within the short-term forecast (up to 120h)
- Workflows exist. Improvements, e.g. in error prediction or members summation, can be quite easily applied
- As a side product: almost 2-yearly verification of several fields of ensemble building models (with relatively easy automation to be possibly implemented)

Drawbacks:

- whatever the change in constituting models (new model, new model version), a yearly verification of new model should be performed

Thank you for your attention

Marta Kopeć[3] - project manager,
Magdalena Gruziel-Słomka[4] - questions about methodology,
Anna Jagodnicka - models verification with `metplus`,
the rest of ICM Meteo Team: Bartosz Niezgódka, Leszek Herman-Iżycki, Małgorzata Melonek, Magdalena Mozga, Antoni Zbytniewski

[3] m.kopec@icm.edu.pl
[4] m.gruziel@icm.edu.pl