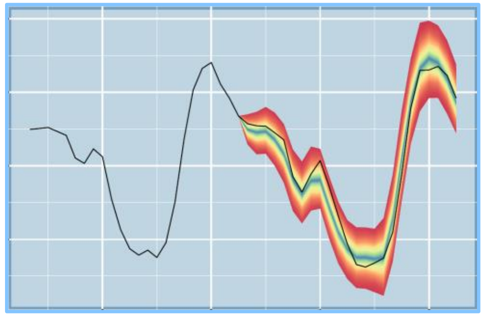
Probabilistic Multi-Step-Ahead Short-Term Water Demand Forecasting with Lasso



Speaker: Jens Kley-Holsteg

GOR Workshop: Entscheidungstheorie und-praxis, OR im Umweltschutz, Projektmanagement und Scheduling

Bochum, 09.10.2020

Authors: Jens Kley-Holsteg, Florian Ziel

UNIVERSITÄT
DUISBURG
ESSEN

Open-Minded

Agenda

Motivation

Application of short-term forecasts in water management and planning.

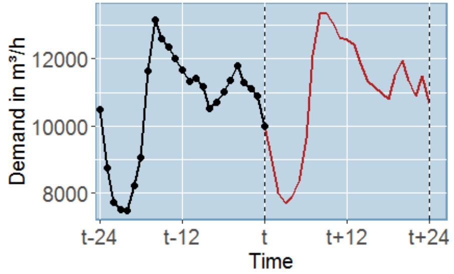
- A popular application of water demand forecasting is the optimized control of **water storage capacities** to
  - reduce energy costs and
  - increase security of supply.

- In this context an appropriate forecasting model should provide:

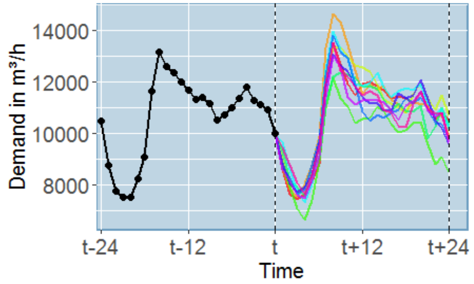
A fast computation

A simple interpretation

A complete representation…



…,which considers beside the mean…



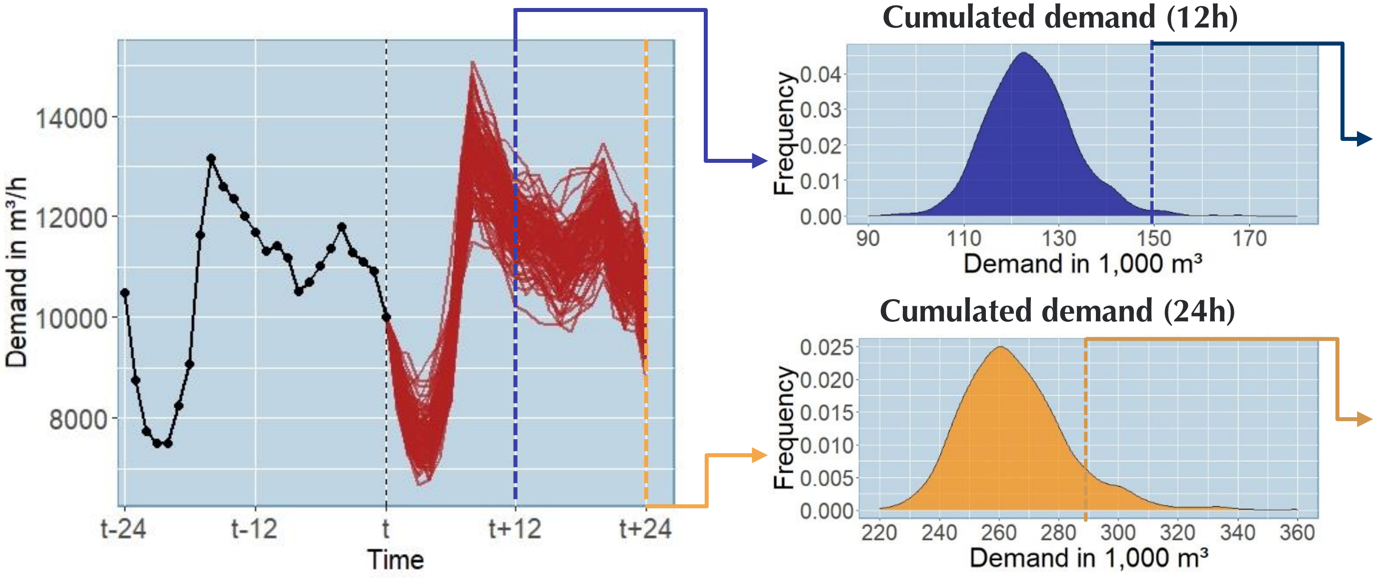
…also the inherent uncertainty.

How can a complete forecasting representation be used to control water storages more efficiently?

Motivation

To optimally control water storages, decision makers are interested in the probability with which a water storage capacity can guarantee the supply over a specific period of time.

Example



Given a **capacity of 150,000 m³** and a **period of 12 h**; the security of supply can be guaranteed with a probability of **0.987**.

Given a **capacity of 290,000 m³** and a **period of 24 h**; the security of supply can be guaranteed with a probability of **0.965**.

For storage optimization problems the *cumulated demand* is the quantity of interest, so that beside the mean and marginal properties the correlation structure within the forecasting horizon must be simulated!

House of Energy Markets & Finance

Agenda

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

Data and Stylized Facts

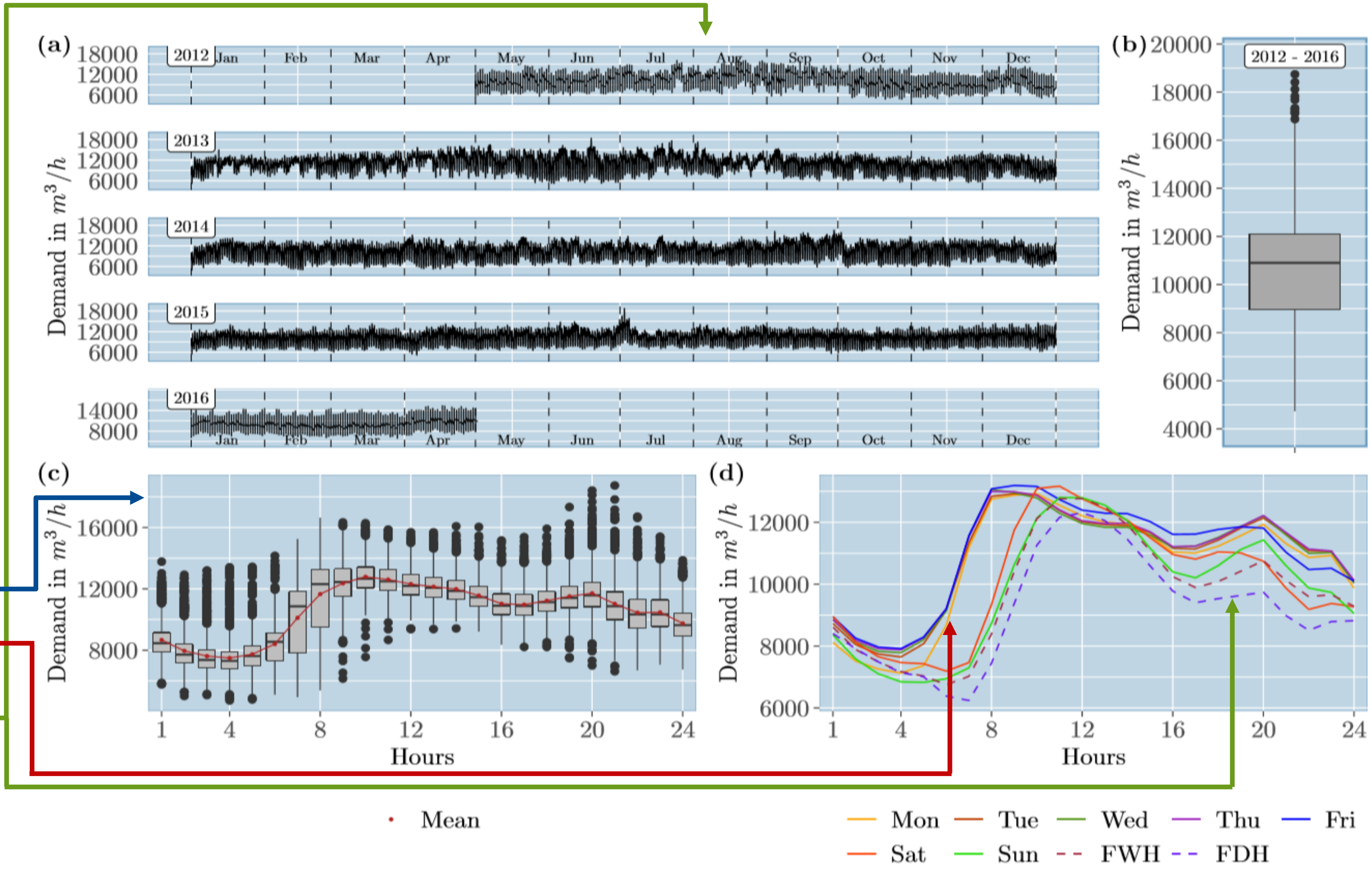
Univariate time series of hourly water demand data in m³/h

(1) Data description and preprocessing

- Six years of hourly data:
  - 4 years for training
  - 2 years for validation
- Data is cleansed:
  - Clock change adjustment
  - NA and measurement error correction (0.1%)

(2) Stylized facts:

- **Daily cycle with varying mean and variance**
- **Weekly cycle with varying weekdays**
- **Yearly cycle with holiday and meteorological effects**
- Stochastic nature and variability



Agenda

Forecasting Model and Estimation Method

Application of high-dimensional linear time series model

To capture the rather complex non-stationary structure of the water demand (Y_t) process, *non-linear models* with a *low-dimensional feature space* are preferred in literature.

However, we choose a rather different approach and introduce a **linear time series model** with a **high-dimensional** feature space:

Feature space exceeds 1,400 features!

Deterministic component: e.g. periodicities as daily, weekly cycle, holidays and meteorological effects.

Time-varying autoregressive comp.: Interaction term between autoregressive and deterministic component.

$$Y_t = \boxed{\mu(t)} + \boxed{\Phi(Y_{t-k})} + \boxed{\Psi(t, Y_{t-k})} + \boxed{\varepsilon(t)}$$

Time-varying zero mean noise process: To model stochastic nature.

Autoregressive component: Lags of most recent history; e.g. Lag 1,2,24,25…

To obtain a parsimonious and fast computable forecasting model the huge feature space must be efficiently tuned!

Forecasting Model and Estimation Method

Estimation method in a linear framework

As estimation method, the **l**east **a**bsolute **s**hrinkage and **s**election **o**perator (**lasso**) is applied with the **B**ayesian **I**nformation **C**riterion (**BIC**) as selection criterion :

$$\widetilde{\hat{\beta}}_{\lambda}^{lasso} = \min_{\beta} || \, \widetilde{Y} - \beta^{T} \, \widetilde{X} \, ||_2^2 + \lambda||\beta||_1,$$

Where:

$\widetilde{Y}$: scaled water demand vector
$\widetilde{X}$: scaled predictor matrix
$\widetilde{\hat{\beta}}$: scaled estimated lasso coefficients
λ : tuning parameter

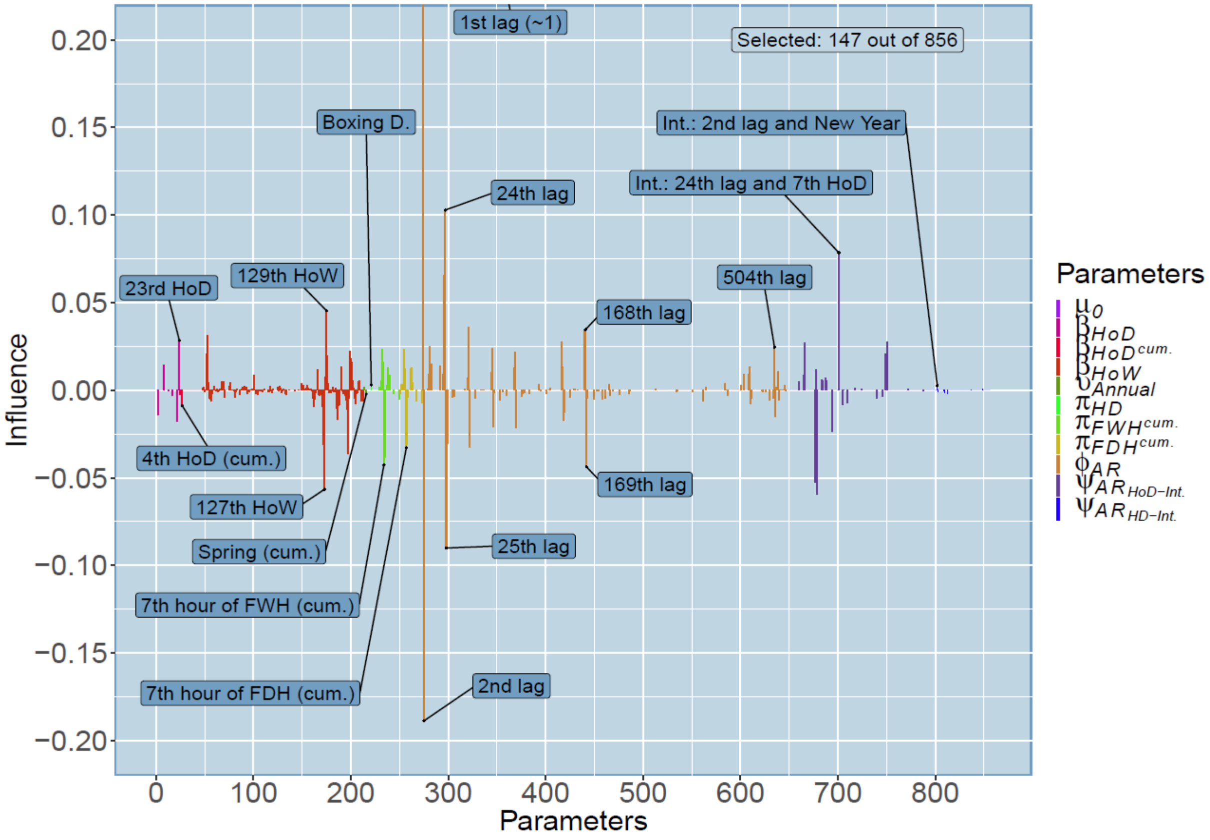
The lasso estimator stands out as…

1. …less important features lose on influence.
2. …irrelevant features are fully excluded.
3. …it is fast computable.
4. …it is simple interpretable.

Accordingly to the lasso algorithm, which features are considered as most influential?

Forecasting Model and Estimation Method

Variable importance of conditional mean estimation



Conditional mean estimation

- For the conditional mean estimation 147 out of 856 features are considered as relevant.
- Most influential features are:
  - Lag 1,2, 24, 25, 168,169…
  - Interaction between lag 24 and hour 7 of the day
  - Hour 127 and 129 of the week
  - Hour 7 of fixed date holidays (cum.)
- It is striking, that each component includes influential features.

Variable importance underlines the necessity for applying a huge feature space!

Agenda

Forecasting framework – Modelling of Uncertainty

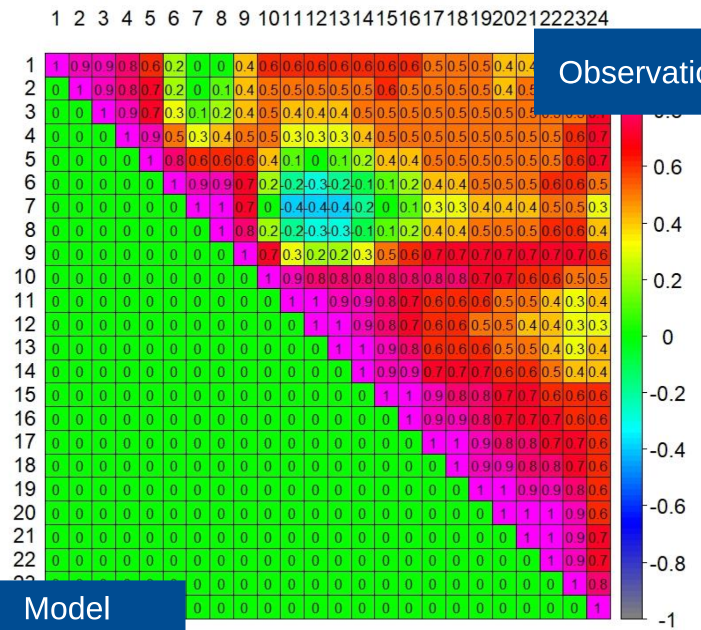
UNIVERSITÄT DUISBURG ESSEN
Open-Minded

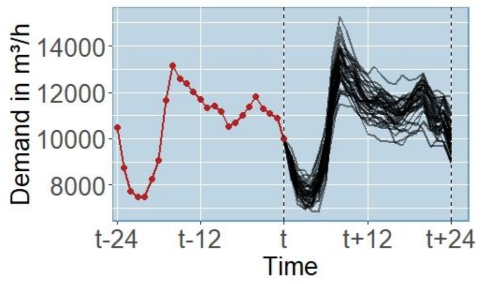
Modelling the prediction uncertainty with a focal point on the cross-correlation structure

To model the **prediction uncertainty** the presented forecasting model is **recursively** solved in a Monte Carlo Simulation Study with sample size of M = 1,000.

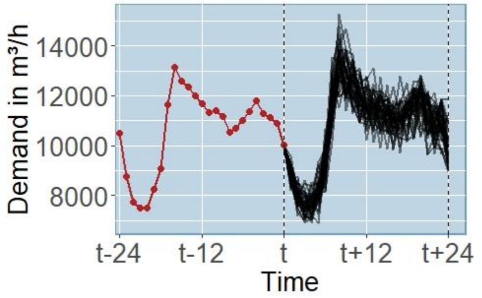
Manipulated Approach

Independent Model Simulation



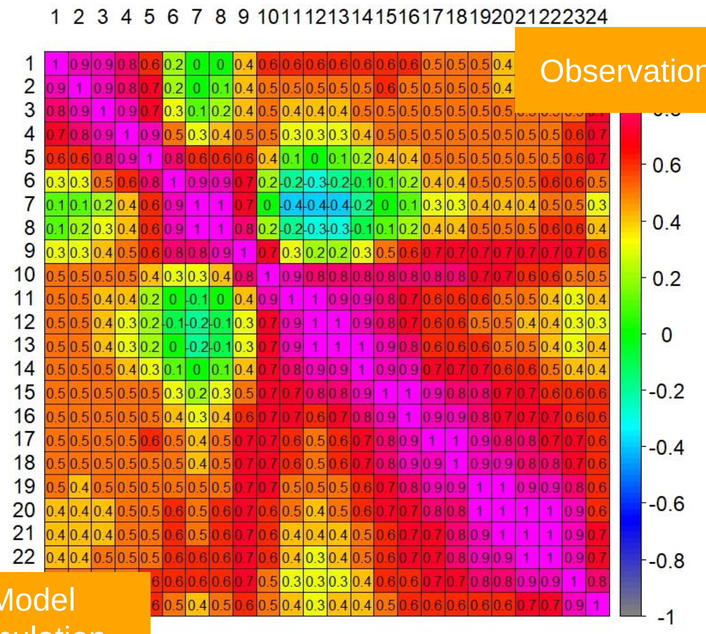


Mean and marginal properties are identical!



Standard Approach

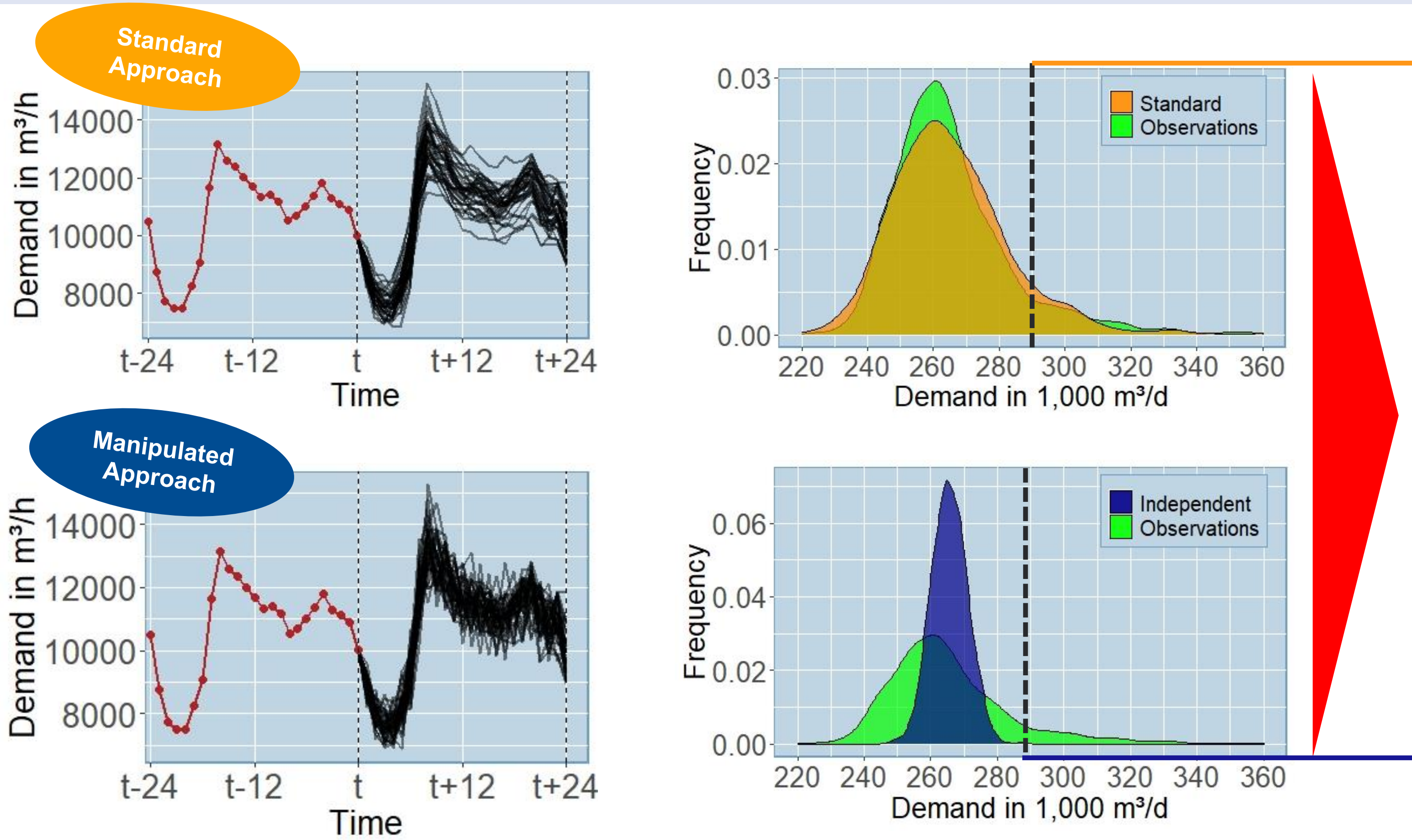
Standard Model Simulation



Crucial is the appropriate simulation of path-dependencies beside marginal properties to quantify the inherent uncertainty of the water demand process!

Forecasting framework – Modelling of Uncertainty

The importance of modelling the path-dependency correctly for water storage optimization



Given a **capacity of 290,000 m³** and a **period of 24 h**; the security of supply can be guaranteed with a probability of:

True	Standard	Manipulated
0.9066	**0.9065**	0.9999

Crucial is the appropriate simulation of path-dependencies beside marginal properties to quantify the inherent uncertainty of the water demand process!

Agenda

Evaluation

How to evaluate probabilistic forecasts?

By now water demand forecasting was focused on **point forecasting** so that in literature mainly point forecasting evaluation measures as the **MAE** and **RMSE** are used!

As we are staying in a probabilistic multi-step-ahead forecasting framework those measures are not sufficient anymore. Hence, the energy score is introduced as an appropriate evaluation measure:

- It is a "strictly proper" scoring rule, so that only the perfect forecast minimizes the criterion
- It allows to discriminate errors in mean, variance and **correlation**
- It generalizes the well-known CRPS for multivariate forecasts H > 1
- It is applicable to distribution forecasts and the implementation is straight forward

Hence, the energy score is the determining performance measure (out of sample!) in the following performance evaluation!

Agenda

Forecasting Performance Evaluation

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

Out-of-sample data

Models from literature

Proposed forecasting model with manipulated correlation structure

Proposed forecasting model

Energy Score as determining measure

Considered as benchmark for "Imp. (%)"-computation

Models	ES (m³/h)	Imp. (%)	PB (m³/h)	Imp. (%)	MAE (m³/h)	Imp. (%)	RMSE (m³/h)	Imp. (%)	NS (%)	Imp. (%)
$Naive_{Mean}$	3,766.82	−92.92	321.16	−98.58	881.77	−104.30	1,056.75	−95.75	69.65	−23.62
$Naive_{FM}$	3,173.94	−62.56	282.89	−74.93	769.01	−78.18	851.77	−57.78	77.98	−14.49
$SARIMA(0,1,4)(0,1,1)_{24}$	2,880.88	−47.55	235.15	−45.40	631.48	−46.31	817.87	−51.50	82.09	−9.98
$Naive_{MRW}$	2,775.66	−42.16	228.15	−41.08	609.12	−41.13	777.25	−43.98	81.14	−11.02
ANN_{Her}	2,693.59	−37.96	220.64	−36.43	600.14	−39.05	754.47	−39.76	83.06	−8.91
SVM_{Her}	2,602.74	−33.30	216.90	−34.12	579.92	−34.37	717.37	−32.89	84.83	−6.98
RF_{Her}	2,397.10	−22.77	196.66	−21.60	524.05	−21.42	659.81	−22.22	87.49	−4.06
$SARIMA(0,1,4)(0,1,1)_{168}$	2,272.23	−16.37	189.38	−17.10	508.89	−17.91	633.23	−17.30	87.38	−4.18
ANN_{Pac}	2,025.14	−3.72	167.61	−3.64	447.96	−3.79	553.90	−2.60	90.72	−0.51
$ARXARCHX_{lasso}^{*}$	1,990.54	−1.95	**148.83**	**7.97**	**402.16**	**6.82**	**500.01**	**7.38**	**92.51**	**1.45**
$AR(p)^{D}$	1,983.53	−1.59	164.09	−1.46	437.14	−1.28	548.88	−1.67	90.90	−0.31
$AR(p)^{W}$	1,952.51	0.00	161.72	0.00	431.60	0.00	539.84	0.00	91.19	0.00
$ARXARCHX_{lasso}^{**}$	1,896.58	2.86	**148.83**	**7.97**	**402.16**	**6.82**	**500.01**	**7.38**	**92.51**	**1.45**
$ARXARCHX_{lasso}^{***}$	1,808.32	7.39	**148.83**	**7.97**	**402.16**	**6.82**	**500.01**	**7.38**	**92.51**	**1.45**
$ARXARCHX_{lasso}$	**1,780.13**	**8.83**	**148.83**	**7.97**	**402.16**	**6.82**	**500.01**	**7.38**	**92.51**	**1.45**

Note: Hypothesis of the DM test, that the loss differential series between best (bold values) and second best ranked model is zero, could be rejected at the 0.001 significance level for each considered evaluation criterion.

*With countermonotone model simulations.

**With comonotone model simulations.

***With independent model simulations.

As shown the proposed $ARXARCHX_{lasso}$ dominates all benchmark models significantly.

Agenda

Conclusion and Further Steps

- The proposed ARX-ARCHX$_{lasso}$ model convinces with a high forecasting performance and…
  - is fast computable
  - is easy interpretable and
  - provides a complete representation of the water demand process

- The need for a complete representation, which considers not only the mean but also the marginal properties and the correlation structure could be highlighted.
  - Remember that the quantity of interest in storage optimization the cumulated water demand can only be computed based on a probabilistic forecast, which considers beside marginal properties also the path-dependencies within the forecasting horizon!

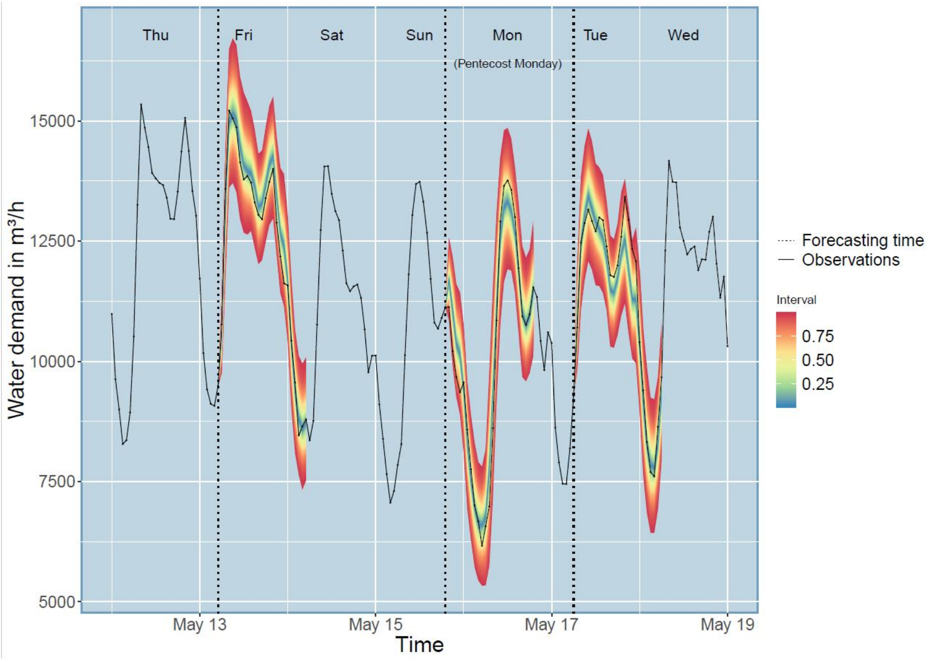
- It could be shown that linear models with a high-dimensional feature space dominate non-linear models with a low-dimensional feature space.
  - Here, it is worth noting, that the amount of Information provided to the model and the ability to efficiently handle this information is the decisive factor.

Thank you for your attention!

Jens Kley-Holsteg
E-mail: jens.kley-holsteg@stud.uni-due.de

Authors: Florian Ziel and Jens Kley-Holsteg

University of Duisburg-Essen



Refrences

- Alvisi, S., and M. Franchini. 2017. “Assessment of predictive uncertainty within the framework of water demand forecasting using the model conditional processor (MCP).” Urban Water J. 14 (1): 1–10. https://doi.org/10.1080/1573062X.2015.1057182.
- Anele, A., Y. Hamam, A. Abu-Mahfouz, and E. Todini. 2017. “Overview, comparative assessment and recommendations of forecasting models for short-term water demand prediction.” Water 9 (11): 887. https://doi.org/10.3390/w9110887.
- Arandia, E., A. Ba, B. Eck, and S. McKenna. 2016. “Tailoring seasonal time series models to forecast short-term water demand.” J. Water Resour. Plann. Manage. 142 (3): 04015067. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000591.
- Bakker, M., H. van Duist, K. van Schagen, J. Vreeburg, and L. Rietveld.2014. “Improving the performance of water demand forecasting modelsby using weather input.” Procedia Eng. 70: 93–102. https://doi.org/10.1016/j.proeng.2014.02.012.
- Bata, M. H., R. Carriveau, and D. S.-K. Ting. 2020. “Short-term water demand forecasting using nonlinear autoregressive artificial neural networks.”J. Water Resour. Plann. Manage. 146 (3): 04020008. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001165.
- Diebold, F. X., and R. S. Mariano. 1995. “Comparing predictive accuracy.” J. Bus. Econ. Stat. 13 (3): 253–263. https://doi.org/10.1198/073500102753410444.
- Donkor, E. A., T. A. Mazzuchi, R. Soyer, and J. Alan Roberson. 2014. “Urban water demand forecasting: Review of methods and models.” J. Water Resour. Plann. Manage. 140 (2): 146–159. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000314.
- Franses, P. H. 2016. “A note on the mean absolute scaled error.” Int. J. Forecasting 32 (1): 20–22. https://doi.org/10.1016/j.ijforecast.2015 .03.008.
- Friedman, J., T. Hastie, and R. Tibshirani. 2010. “Regularization paths for generalized linear models via coordinate descent.” J. Stat. Software 33 (1): 1. https://doi.org/10.18637/jss.v033.i01.
- Gagliardi, F., S. Alvisi, M. Franchini, and M. Guidorzi. 2017a. “A comparison between pattern-based and neural network short-term water demand forecasting models.” Water Sci. Technol. Water Supply 17 (5): 1426–1435. https://doi.org/10.2166/ws.2017.045.

Refrences

- Gagliardi, F., S. Alvisi, Z. Kapelan, and M. Franchini. 2017b. “A probabilistic short-term water demand forecasting model based on the Markov chain.” Water 9 (7): 507. https://doi.org/10.3390/w9070507.
- Gneiting, T., and M. Katzfuss. 2014. “Probabilistic forecasting.” Annu. Rev. Stat. Appl. 1 (1): 125–151. https://doi.org/10.1146/annurev-statistics-062713-085831.
- Gneiting, T., and A. E. Raftery. 2007. “Strictly proper scoring rules, prediction, and estimation.” J. Am. Stat. Assoc. 102 (477): 359–378. https://doi.org/10.1198/016214506000001437.
- Gneiting, T., L. I. Stanberry, E. P. Grimit, L. Held, and N. A. Johnson. 2008. “Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds.” TEST 17 (2): 211–235. https://doi.org/10.1007/s11749-008-0114-x.
- Hastie, T., M. Wainwright, and R. Tibshirani. 2015. “Statistical learning with Sparsity: The lasso and generalizations.” In Monographs on statistics and applied probability. Boca Raton, FL: CRC Press.
- Herrera, M., L. Torgo, J. Izquierdo, and R. Pérez-García. 2010. “Predictive models for forecasting hourly urban water demand.” J. Hydrol. 387 (1–2): 141–150. https://doi.org/10.1016/j.jhydrol.2010.04.005.
- Hutton, C. J., and Z. Kapelan. 2015. “A probabilistic methodology for quantifying, diagnosing and reducing model structural and predictive errors in short term water demand forecasting.” Environ. Modell. Software 66 (Apr): 87–97. https://doi.org/10.1016/j.envsoft.2014.12.021.
- Hutton, C. J., Z. Kapelan, L. Vamvakeridou-Lyroudia, and D. A. Savi´c. 2014. “Dealing with uncertainty in water distribution system models: A framework for real-time modeling and data assimilation.” J. Water Resour. Plann. Manage. 140 (2): 169–183. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000325.
- Hyndman, R., G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O’Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, and F. Yasmeen.
- 2019. “Forecast: Forecasting functions for time series and linear models.” Accessed December 13, 2019. http://pkg.robjhyndman.com/forecast.

Refrences

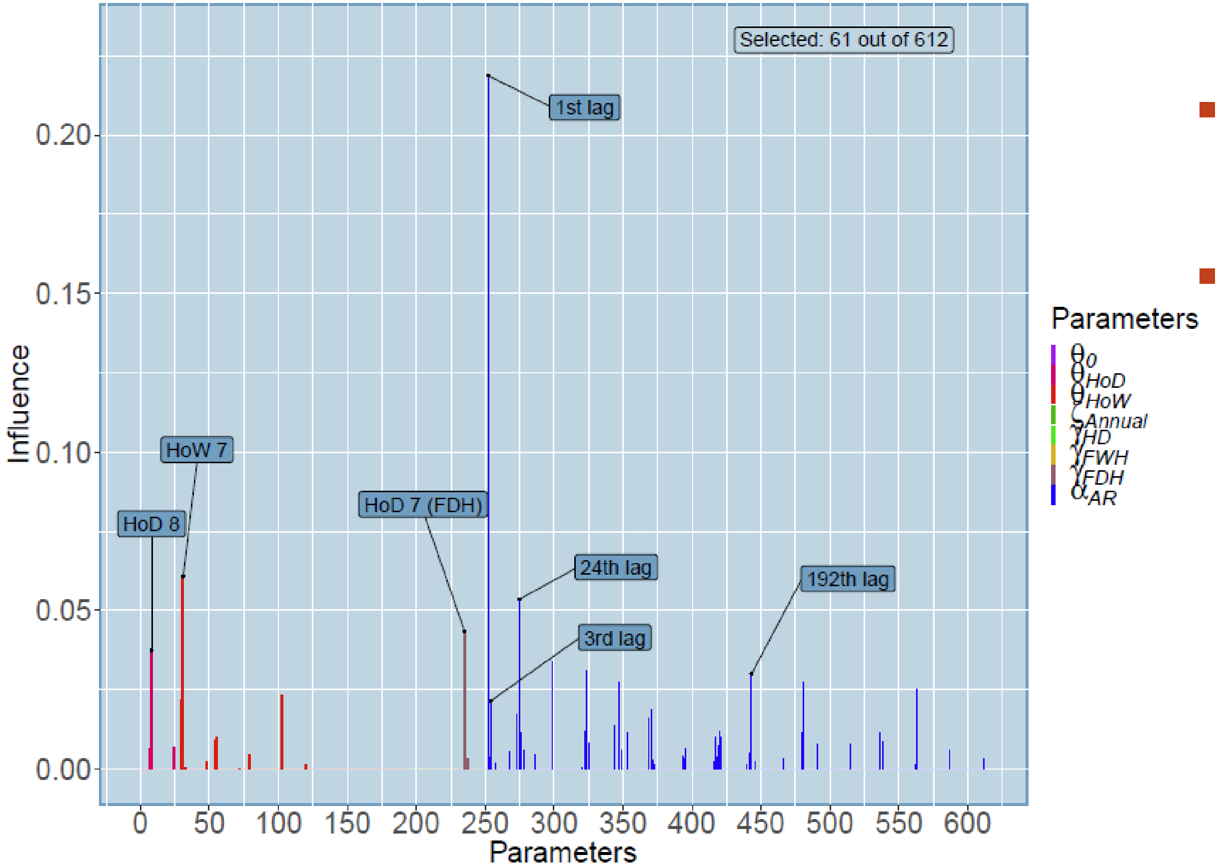
- Hyndman, R. J., and Y. Khandakar. 2008. "Automatic time series forecasting: The forecast package for R." J. Stat. Software 27 (3): 1–22. Liaw, A., and M. Wiener. 2002. "Classification and regression by randomforest." R News 2 (3): 18–22.
- Meyer, D., E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch. 2019. "e1071: Misc functions of the department of statistics, probability theory group (Formerly: E1071)." Accessed December 7, 2019. https://CRAN.R-project.org/package=e1071.
- Neath, A. A., and J. E. Cavanaugh. 2012. "The Bayesian information criterion: Background, derivation, and applications." Wiley Interdiscip. Rev. Comput. Stat. 4 (2): 199–203. https://doi.org/10.1002/wics.199.
- Nowotarski, J., and R. Weron. 2018. "Recent advances in electricity price forecasting: A review of probabilistic forecasting." Renewable Sustainable Energy Rev. 81 (Jan): 1548–1568. https://doi.org/10.1016/j.rser.2017.05.234.
- Pacchin, E., S. Alvisi, and M. Franchini. 2017. "A short-term water demand forecasting model using a moving window on previously observed data." Water 9 (3): 172. https://doi.org/10.3390/w9030172.
- Pacchin, E., F. Gagliardi, S. Alvisi, and M. Franchini. 2019. "A comparison of short-term water demand forecasting models." Water Resour. Manage. 33 (4): 1481–1497. https://doi.org/10.1007/s11269-019-02213-y.
- R Core Team. 2019. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S. New York: Springer.
- Ziel, F. 2016. "Iteratively reweighted adaptive lasso for conditional heteroscedastic time series with applications to AR–ARCH type processes." Comput. Stat. Data Anal. 100 (Aug): 773–793. https://doi.org/10.1016/j.csda.2015.11.016.
- Ziel, F. 2018. "Modeling public holidays in load forecasting: A German case study." J. Mod. Power Syst. Clean Energy 6 (2): 191–207. https://doi.org/10.1007/s40565-018-0385-5.

Refrences

- Ziel, F., C. Croonenbroeck, and D. Ambach. 2016. "Forecasting wind power: Modeling periodic and non-linear effects under conditional heteroscedasticity." Appl. Energy 177 (Sep): 285–297. https://doi.org/10.1016/j.apenergy.2016.05.111.
- Ziel, F., and B. Liu. 2016. "Lasso estimation for gefcom2014 probabilistic electric load forecasting." Int. J. Forecasting 32 (3): 1029–1037. https://doi.org/10.1016/j.ijforecast.2016.01.001.
- Ziel, F., R. Steinert, and S. Husmann. 2015. "Efficient modeling and forecasting of electricity spot prices." Energy Econ. 47 (Jan): 98–111. https://doi.org/10.1016/j.eneco.2014.10.012.

Forecasting Model and Estimation Method

Variable importance of conditional variance estimation



Conditional variance estimation

- For the conditional variance estimation 61 out of 612 features are considered as relevant.
- Most influential features are:
  - Lag 1,3, 24, 48,…
  - Hour 8 of the day
  - Hour 7 of the week
  - Hour 7 of a fixed date holiday

For the conditional mean as well as conditional variance estimation the lasso estimator allows to reduce efficiently the feature space!

Forecasting framework – Modelling of Uncertainty

UNIVERSITÄT DUISBURG ESSEN

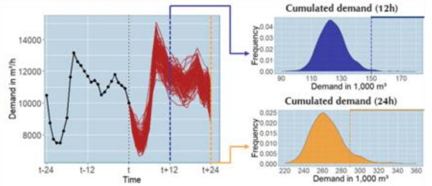
Open-Minded

To provide a complete forecast, the inherent uncertainty of the water demand process must be modelled

As already noted, the water demand process is stochastic in nature, so that the forecasting model must be able to quantify and issue the true variability - called **predication uncertainty**.

In a multi-step-ahead forecasting framework increases the complexity significantly as not only the marginal properties but also the **path-dependency** within the forecasting horizon are of interest.

Remember: To quantify the expected cumulated demand, we rely on the evolution of each sample path over time (path-dependency)!



Not quantity of interest for practitioners!

The **emulation uncertainty** arising and cascading within the data collection and modelling procedure is not the quantity of interest and must be quantified **but** marginalized, so that the probabilistic forecaster issues only the natural variability!

e.g.

- Measurement/ data uncertainty
- Parameter uncertainty
- Model structure uncertainty

Hence, we are concerned to issue only the prediction uncertainty. But how can the prediction uncertainty be modelled so that beside marginal properties also the correlation structure within the forecasting horizon is considered?

Forecasting Performance Evaluation

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

In-sample data

Compared to typical ML-Alg. is the number of parameters of the proposed model rather low

In-sample only point forecasting measures are applicable

Models from literature

Proposed forecasting model

Considered as benchmark for "Imp. (%)"-computation

Caused by overfitting

Models	Data length (h)	Parameters (active)	Parameters (possible)	MAE[a] (m^3/h)	Imp. (%)	RMSE[a] (m^3/h)	Imp. (%)	NS[a] (%)	Imp. (%)
$Naive_{FM}$	35,064	24	24	888.06	−293.12	1,175.64	−269.25	68.10	−30.27
$Naive_{Mean}$	35,064	144	144	731.92	−224.00	967.78	−203.96	78.35	−19.78
$Naive_{MRW}$	35,064	0	0	644.54	−185.32	947.52	−197.60	79.29	−18.81
ANN_{Pac}	8,760	3,624	3,624	474.26	−109.94	665.60	−109.06	89.41	−8.45
RF_{Her}	1,344	782,897	>782,897	294.67	−30.44	405.83	−27.46	96.11	−1.59
$SARIMA(0,1,4)(0,1,1)_{24}$	672	6	6	288.79	−27.84	404.71	−27.11	96.11	−1.59
SVM_{Her}	1,344	630	>630	278.16	−23.13	373.87	−17.43	96.69	−0.99
ANN_{Her}	1,344	71	71	262.24	−16.08	344.06	−8.07	97.20	−0.47
$AR(p)^D$	35,064	1,280	1,668	230.60	−2.08	320.02	−0.51	97.63	−0.03
$AR(p)^W$	35,064	401	1,668	225.90	0.00	318.39	0.00	97.66	0.00
$ARXARCHX_{lasso}$	35,064	200	1,468	216.13	4.33	300.16	5.73	97.91	0.26
$SARIMA(0,1,4)(0,1,1)_{168}$	672	6	6	**185.20**	**18.02**	**291.82**	**8.34**	**97.93**	**0.28**

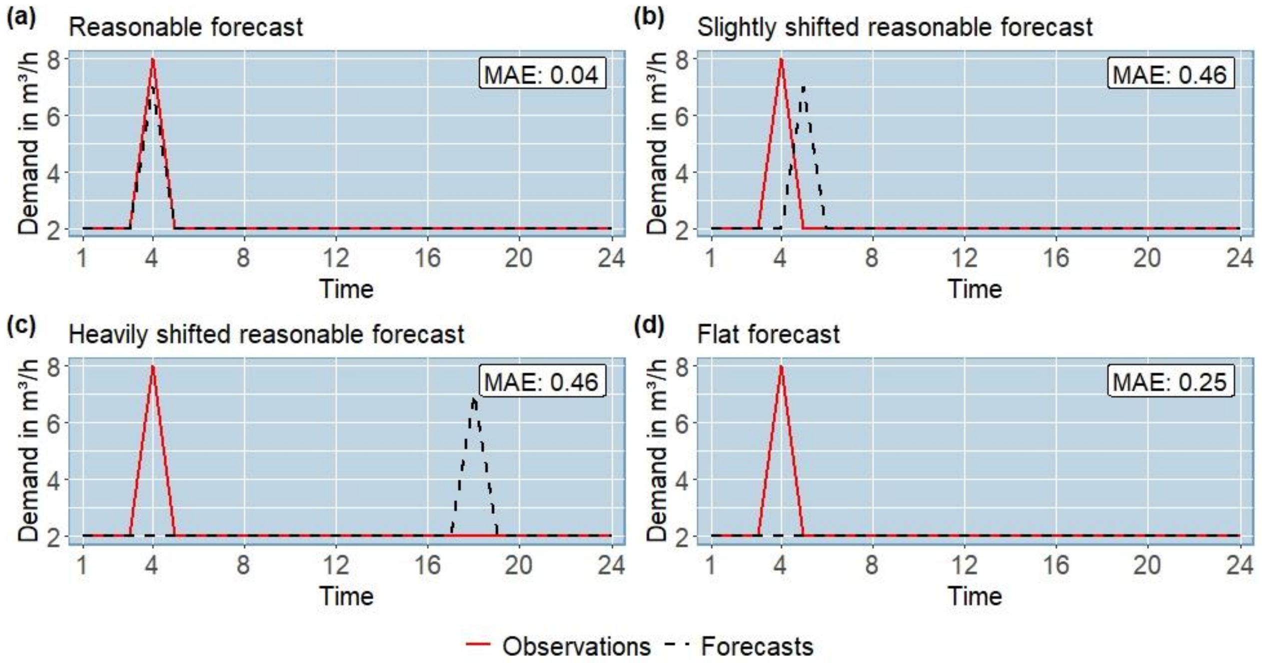
Note: Hypothesis of the DM test, that the loss differential series between best (bold values) and second best ranked model is zero, could be rejected at the 0.001 significance level for each considered evaluation criterion.

[a]Within the calibration period the forecasting horizon H in Eq. (14) (MAE), Eq. (15) (RMSE), and Eq. (16) (NS) is set equal to 1.

As shown the proposed $ARXARCHX_{lasso}$ is a parsimonious and simple interpretable model compared to existing ML-Alg.. Moreover it stands out with a high in-sample forecasting accuracy.

Evaluation

As we are in a probabilistic multi-step-ahead forecasting framework, existing point forecasting measures as the MAE and RMSE are not sufficient anymore!



Example: Shortcoming of MAE

- In terms of storage optimization, forecast (b) is considered as moderate.
- However, in terms of the MAE forecast (b) achieves with forecast (c) the worst score.

The MAE is not able to assess multiple time steps at once!

Hence, a measures is required, which penalizes simultaneously errors in the mean, the marginal properties and the correlation structure!

Scoring rules provide an aggregated measure by assigning a numerical score based on the issued distribution F and the events y that materialize.

- What makes the energy score so appealing as evaluation criteria?
  - It is a "strictly proper" scoring rule, so that only the perfect forecast minimizes the criterion
  - It allows to discriminate errors in mean, variance and **correlation**
  - It generalizes the well-known CRPS for multivariate forecasts H > 1
  - It is applicable to distribution forecasts and the implementation is straight forward

- Energy Score:

$$ES_\beta(\boldsymbol{F_X}, \boldsymbol{y}) = \mathbb{E}\left[||\boldsymbol{X} - \boldsymbol{y}||_2^\beta\right] - \frac{1}{2}\mathbb{E}\left[||\boldsymbol{X} - \tilde{\boldsymbol{X}}||_2^\beta\right], \text{ whereby, } \beta = 1; \boldsymbol{X}, \tilde{\boldsymbol{X}} \sim \boldsymbol{F_X}$$