universität innsbruck

Graphical Assessment of Probabilistic Precipitation Forecasts

Reto Stauffer, Moritz N. Lang, Achim Zeileis

https://topmodels.R-Forge.R-project.org/

Introduction

Probabilistic predictions

- Modelling full probabilistic distribution
- Allows to retrieve the expected value, probabilities, exceedances, . . .
- Important in many fields (e.g., medicine, economics, meteorology, . . .)

Introduction

Probabilistic predictions

- Modelling full probabilistic distribution
- Allows to retrieve the expected value, probabilities, exceedances, . . .
- Important in many fields (e.g., medicine, economics, meteorology, . . .)

Objective

- Increasing sharpness conditional on calibration (*Gneiting et al. 2007a*)
- Optimization/model selection: proper scoring rules (*Gneiting et al. 2007b*)
- Graphical assessment: goodness of fit and possible misspecification

Case study

Probabilistic precipitation forecasting:

Accurate and reliable precipitation forecasts of increasing importance for e.g.:

- Tourism
- Agricultural applications
- Road safety and maintenance during winter season
- Risk assessment (droughts, floods, fire hazard, . . .)
- Strategic resource planning (water supply, hydro power, transport, . . .)

Case study

Probabilistic precipitation forecasting:

Accurate and reliable precipitation forecasts of increasing importance for e.g.:

- Tourism
- Agricultural applications
- Road safety and maintenance during winter season
- Risk assessment (droughts, floods, fire hazard, . . .)
- Strategic resource planning (water supply, hydro power, transport, . . .)

$\Rightarrow$ **Statistical weather prediction 'detour'**

Case study

Weather forecasts

- Typically physically-based numerical weather prediction models
- Multiple runs with modified conditions $\rightarrow$ ensemble forecasts
- Various sources of possible errors due to necessary simplifications

Case study

Weather forecasts

- Typically physically-based numerical weather prediction models
- Multiple runs with modified conditions $\rightarrow$ ensemble forecasts
- Various sources of possible errors due to necessary simplifications

Statistical post-processing

- Use historical observations and ensemble forecasts
- Estimate statistical models to correct for possible forecast errors in both, expectation and uncertainty
- Apply correction to latest ensemble forecast

Case study

Data

- Station Innsbruck Airport, Austria
- 13 years of daily records ($2000 - 2013$; $N = 4971$)

Case study

Data

- Station Innsbruck Airport, Austria
- 13 years of daily records ($2000 - 2013$; $N = 4971$)
- Response: Observed 3 day accumulated precipitation (`rain`)
- Features: mean and standard deviation of accumulated precipitation (11-member ensemble; $5 - 8$ days ahead; `ensmean`, `enssd`)

Case study

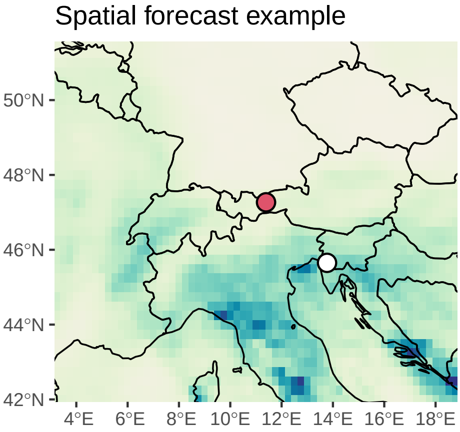
Data

- Station Innsbruck Airport, Austria
- 13 years of daily records ($2000 - 2013$; $N = 4971$)
- Response: Observed 3 day accumulated precipitation (`rain`)
- Features: mean and standard deviation of accumulated precipitation (11-member ensemble; $5 - 8$ days ahead; `ensmean`, `enssd`)

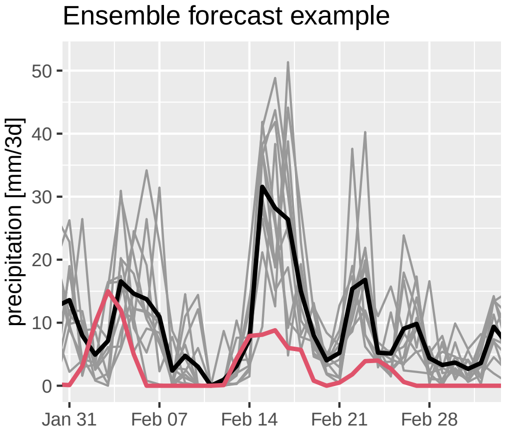
Study goal

- Estimate three different parametric regression models
- Assessing goodness of fit using graphical assessment methods

Case study

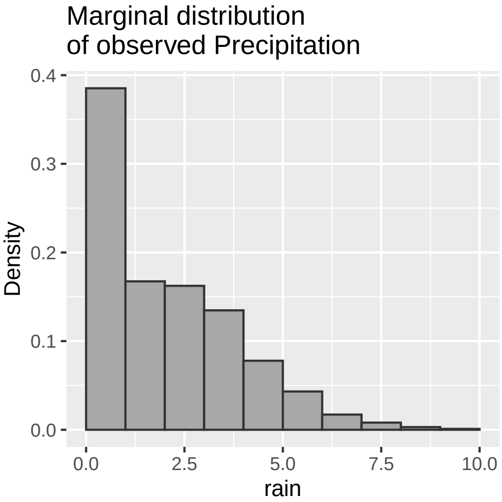


Spatial forecast example
50°N
48°N
46°N
44°N
42°N
4°E
6°E
8°E
10°E
12°E
14°E
16°E
18°E

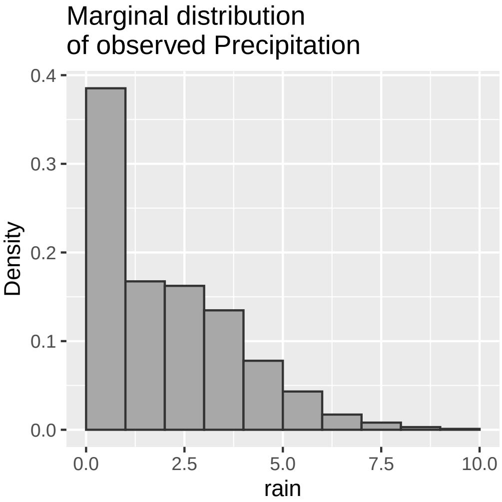


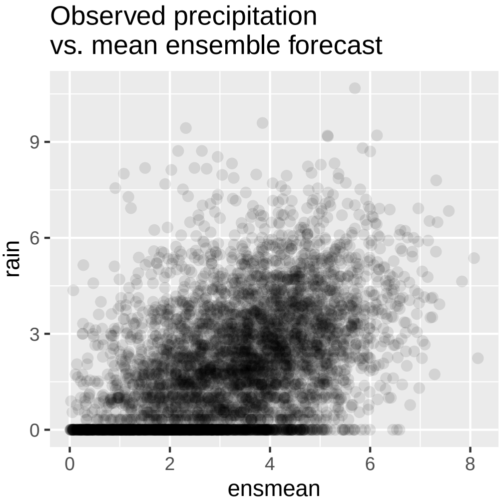
Ensemble forecast example
precipitation [mm/3d]
50
40
30
20
10
0
Jan 31
Feb 07
Feb 14
Feb 21
Feb 28

Use case


Marginal distribution
of observed Precipitation
Density
0.4
0.3
0.2
0.1
0.0
0.0
2.5
5.0
7.5
10.0
rain

Use case


Marginal distribution
of observed Precipitation
Density
0.4
0.3
0.2
0.1
0.0
0.0
2.5
5.0
7.5
10.0
rain


Observed precipitation
vs. mean ensemble forecast
rain
9
6
3
0
0
2
4
6
8
ensmean

Weather Forecasting

Statistical models:

Revisiting models by Messner, Mayr, and Zeileis (2010):

	Distribution	Location	Scale
ols	$y_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$	$\hat{\mu}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot \text{ensmean}_i$	$\log(\hat{\sigma}_i) = \hat{\gamma}_0$

Weather Forecasting

Statistical models:
Revisiting models by Messner, Mayr, and Zeileis (2010):

	Distribution	Location	Scale
ols	$y_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$	$\hat{\mu}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot \text{ensmean}_i$	$\log(\hat{\sigma}_i) = \hat{\gamma}_0$
hcnorm	$y_i \sim \mathcal{N}_0(\mu_i, \sigma_i^2)$	$\hat{\mu}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot \text{ensmean}_i$	$\log(\hat{\sigma}_i) = \hat{\gamma}_0 + \hat{\gamma}_1 \cdot \log(\text{enssd}_i)$
hclog	$y_i \sim \mathcal{L}_0(\mu_i, \sigma_i^2)$	$\hat{\mu}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot \text{ensmean}_i$	$\log(\hat{\sigma}_i) = \hat{\gamma}_0 + \hat{\gamma}_1 \cdot \log(\text{enssd}_i)$

Model assessment

Scores: Continuous ranked probability score (CRPS) and logScore:

	ols	hcnorm	hclog
CRPS	0.913	0.877	0.876
logScore	1.915	1.804	1.799

Model assessment

Scores: Continuous ranked probability score (CRPS) and logScore:

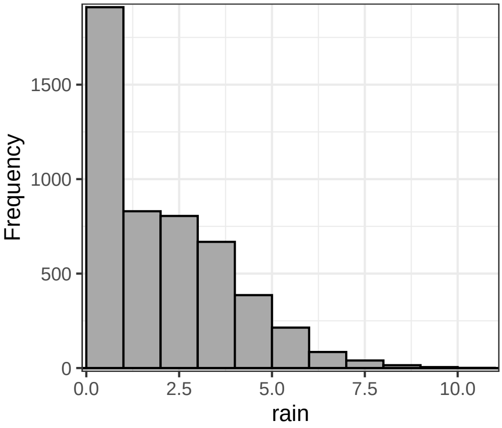
	ols	hcnorm	hclog
CRPS	0.913	0.877	0.876
logScore	1.915	1.804	1.799

Graphical model assessment

- Important complement to proper scoring rules
- Checking marginal and probabilistic calibration
- Allows to identify possible misspecifications

Marginal calibration

Frequencies: Observed

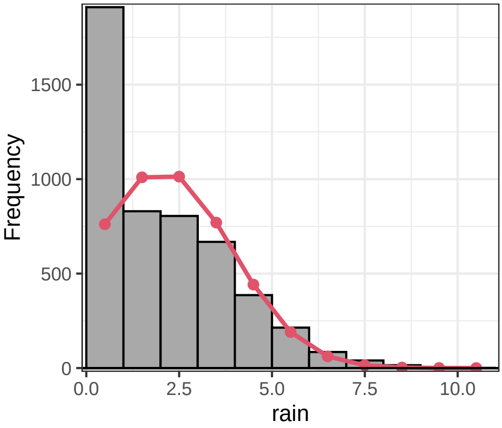


Observed frequency

$$\text{obs}_j = \sum_{i=1}^{N} I\big(y_i \in [b_j, b_{j+1})\big)$$

Marginal calibration

Frequencies: Observed vs. expected



Observed frequency

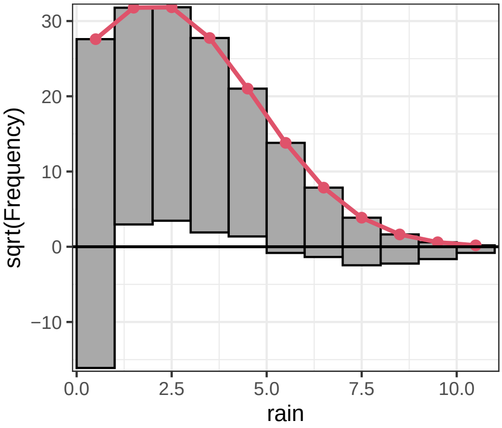
$$\text{obs}_j = \sum_{i=1}^{N} I\big(y_i \in [b_j, b_{j+1})\big)$$

Expected frequency

$$\text{exp}_j = \sum_{i=1}^{N} \big(F(b_{j+1}|\hat{\theta}_i) - F(b_j|\hat{\theta}_i)\big)$$

Marginal calibration

Frequencies: $\sqrt{\text{Observed}}$ vs. $\sqrt{\text{expected}}$



Observed frequency

$$\text{obs}_j = \sum_{i=1}^{N} I\big(y_i \in [b_j, b_{j+1})\big)$$

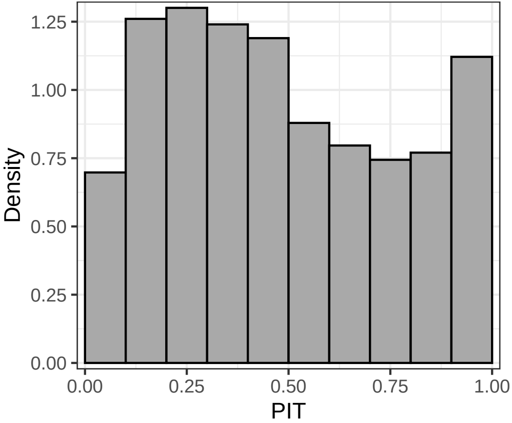
Expected frequency

$$\text{exp}_j = \sum_{i=1}^{N} \big(F(b_{j+1}|\hat{\theta}_i) - F(b_j|\hat{\theta}_i)\big)$$

$\Rightarrow$ **Hanging rootogram**

Probabilistic calibration

PIT residuals



Continuous case

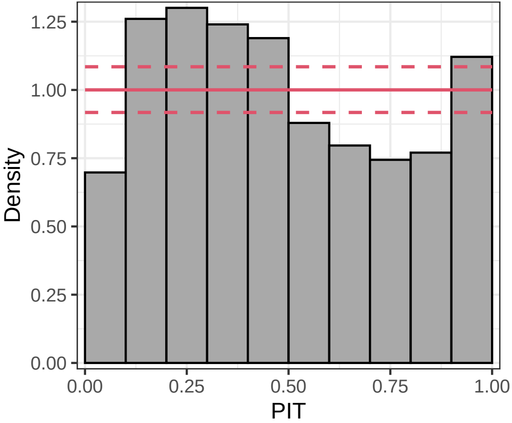
$$u_i = F(y_i|\hat{\theta}_i)$$

Discrete case (*Czado et al. 2009*)

$$u_i = F(y_i - 1|\hat{\theta}_i) + \nu\left[F(y_i - 1|\hat{\theta}_i), F(y_i, |\hat{\theta}_i)\right]$$

Probabilistic calibration

PIT residuals



Continuous case

$$u_i = F(y_i|\hat{\theta}_i)$$

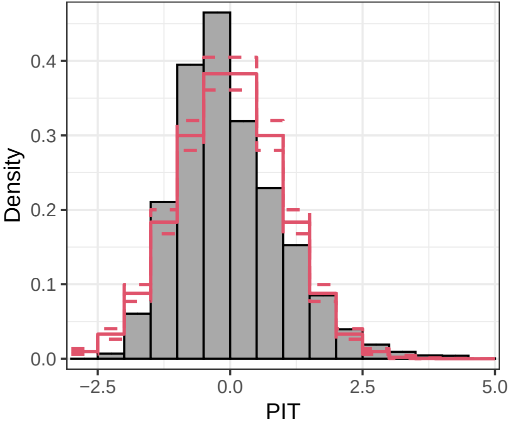
Discrete case (*Czado et al. 2009*)

$$u_i = F(y_i - 1|\hat{\theta}_i) + \nu\left[F(y_i - 1|\hat{\theta}_i), F(y_i, |\hat{\theta}_i)\right]$$

$\Rightarrow$ **Uniform scale: PIT histogram**

Probabilistic calibration

PIT residuals: Normal scale

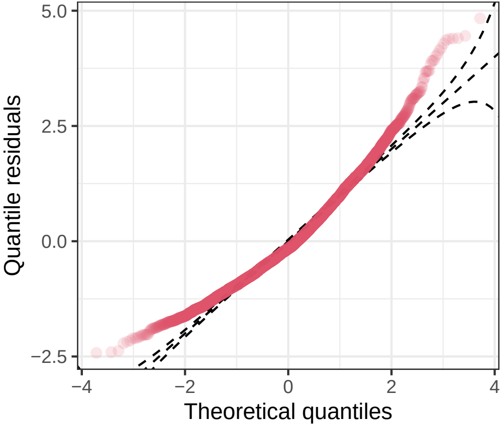


Quantile residuals:

$$\hat{r}_i = \Phi^{-1}\left(F(y_i|\hat{\theta}_i)\right) = \Phi^{-1}(u_i)$$

Probabilistic calibration

Quantile residuals: Observed vs. expected



Quantile residuals:

$$\hat{r}_i = \Phi^{-1}\Big(F(y_i|\hat{\theta}_i)\Big) = \Phi^{-1}(u_i)$$

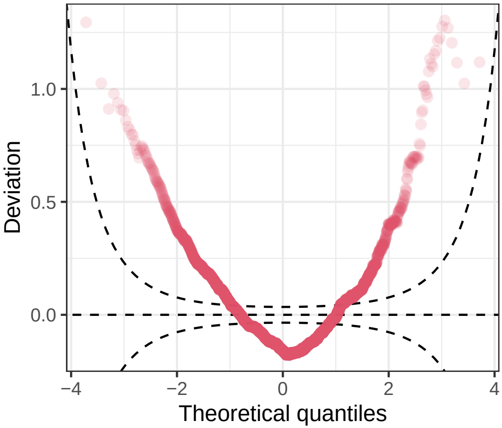
Data pairs:

$$(z_{(1)}, \hat{r}_{(1)}), \ldots, (z_{(N)}, \hat{r}_{(N)})$$

$\Rightarrow$ **(Randomized) Q-Q residual plot**

Probabilistic calibration

Quantile residuals: Deviations



Detrended Q-Q residuals:

$$(z_{(1)}, \hat{r}_{(1)} - z_{(1)}), \ldots, (z_{(N)}, \hat{r}_{(N)} - z_{(N)})$$

$\Rightarrow$ **Wormplot**

topmodels implementation

```
R> library("topmodels")
```

Core functions:

```
R> rootogram(ols)
R> pithist(ols)
R> qqrplot(ols)
R> wormplot(ols)
```

topmodels implementation

```
R> library("topmodels")
```

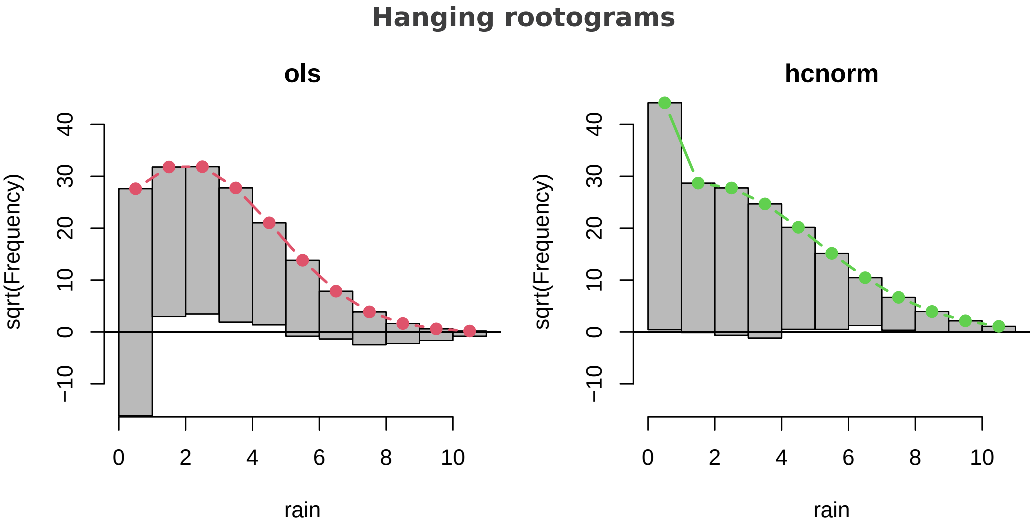
Core functions:

```
R> rootogram(ols)
R> pithist(ols)
R> qqrplot(ols)
R> wormplot(ols)
```

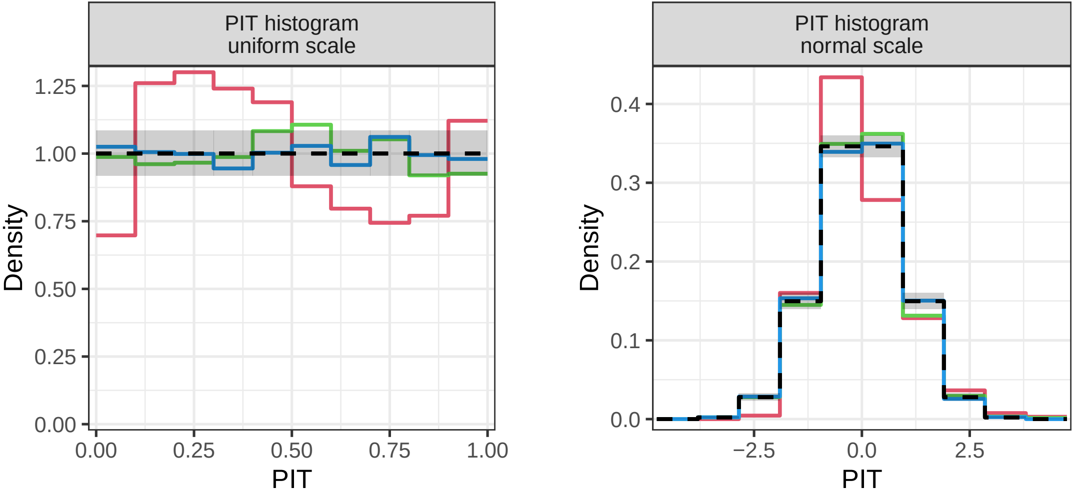
Comparing different models:

```
R> plot(c(pithist(ols), pithist(hcnorm)), ...)
R> plot(c(pithist(ols), pithist(hcnorm)), single_graph = TRUE, style = "l", ...)

R> plot(c(qqrplot(ols), qqrplot(hcnorm)), ...)
R> plot(c(qqrplot(ols), qqrplot(hcnorm)), single_graph = TRUE, ...)

R> plot(c(wormplot(ols), wormplot(hcnorm)), ...)
R> plot(c(wormplot(ols), wormplot(hcnorm)), single_graph = TRUE, ...)
```

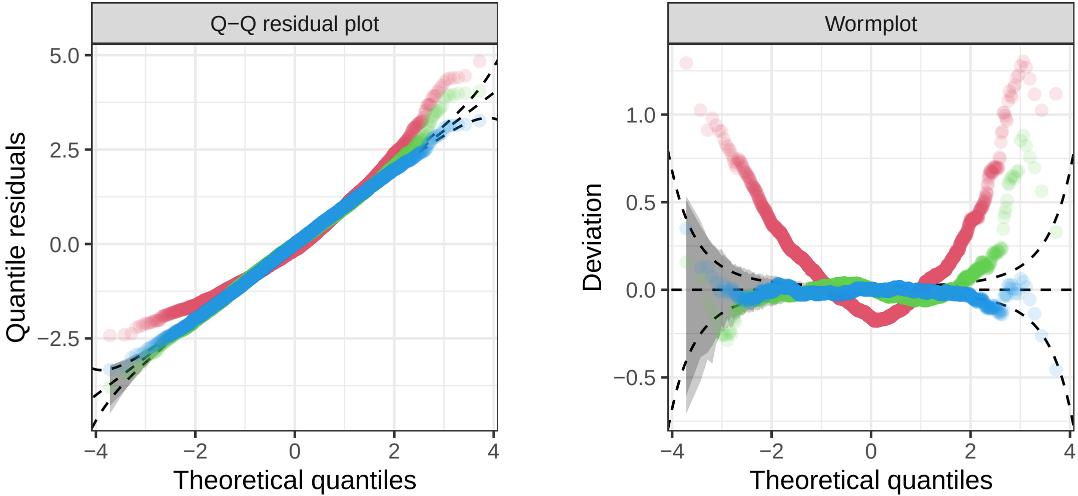
Model comparison



Model comparison



Model comparison



Summary

Graphical assessments:

Various possibilities suggested in different parts of the literature.

- Rootogram
- Probability integral transform (PIT) histogram
- (Randomized) quantile-quantile residuals plot
- Detrended Q-Q residuals plot or worm plot
- Reliability diagram at prespecified thresholds

Summary

topmodels: Unifying toolbox for graphical model assessment.

- available on R-Forge at `https://topmodels.R-Forge.R-project.org/`

Concept: Unifying toolbox for probabilistic forecasts and graphical model assessment.

Graphics: Implemented in R base graphics and `ggplot2`.

Models: `(g)lm`, `crch`, `disttree`, and more to come.

References

Lang MN, Zeileis A *et al.* (2021). "topmodels: Infrastructure for Inference and Forecasting in Probabilistic Models." *R package version 0.2-0*. `https://topmodels.R-Forge.R-project.org/`

Czado C, Gneiting T, Held L (2009). "Predictive Model Assessment for Count Data." *Biometrics*, **65**(4), 1254–1261. `doi:10.1111/j.1541-0420.2009.01191.x`

Dunn PK, Smyth GK (1996). "Randomized Quantile Residuals." *Journal of Computational and Graphical Statistics*, **5**(3), 236–244. `doi:10.2307/1390802`

Gneiting T, Balabdaoui F, Raftery AE (2007a) "Probabilistic Forecasts, Calibration and Sharpness." *Journal of the Royal Statistical Society: Series B (Methodological)*, **69**(2), 243–268. `doi:10.1111/j.1467-9868.2007.00587.x`

Gneiting T, Raftery AE (2007b) "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association*, **102**(477), 359–378. `doi:10.1198/016214506000001437`

Kleiber C, Zeileis A (2016). "Visualizing Count Data Regressions Using Rootograms." *The American Statistician*, **70**(3), 296–303. `doi:10.1080/00031305.2016.1173590`

Messner JW, Mayr GJ, Zeileis A (2016). "Heteroscedastic Censored and Truncated Regression with crch." *The R Journal.*, **8**(1), 173–181. `doi:10.32614/RJ-2016-012`

https://topmodels.R-Forge.R-project.org/

Reto.Stauffer@uibk.ac.at