Environment and Climate Change Canada
Environnement et Changement climatique Canada

Multi-Valued Verification

7th International Verification Methods Workshop
VKU Forum, Berlin
Tim Bullock & Nelson Shum
Meteorological Service of Canada (MSC)
2017-05-09

Contents

- The Case for Multi-valued Verification (MVV)
- Dimensionless Metrics
- Dimensional Metrics
- Applications
- Future work

The Case for Multi-valued Verification

- For objective verification:
  - "It must be clear what is being forecast"
  - "the verification process should…reflect the forecast precisely" (Jolliffe & Stephenson, 2012)
- Many MSC forecasts continuous over space and/or time
  - Examples:
  - Marine wind forecasts
  - Public temperature forecasts
- A few are multi-valued at single point in space and time
  - Example: precipitation type (RASN, FZRAPL, etc.)
- MSC verification methods are single-valued
  - Compare one forecast value with one observed value

The Case for MVV (2)

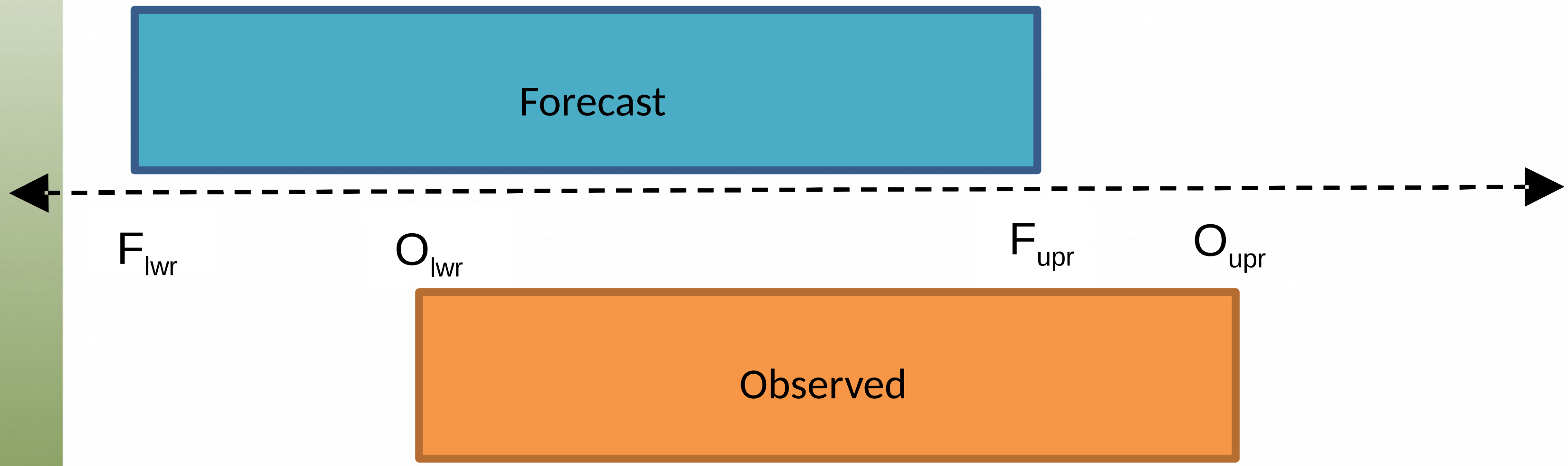
- Summarize to single value(s) for verification
  - "Representative point" assumption
  - Extreme or average or "dominant" value
  - Discretize in time and/or space
- Summarizing can cause information loss
  - Risk of invalidating verification results
  - Forecasters reject results if information losses too great
- Multiple observations often available
  - In situ observing networks
  - Remote sensing (radar, satellite, lightning, etc.)
  - Synthetic observations
- Physical fields usually continuous
  - Exploit this property for verification?

The Case for MVV (3)

- Suppose we allow multiple values
  - For forecasts (F): range or list of values by definition
  - For observations (O): construct range from multiple observations
  - May not cover entire range: “I know this much is true”
- No one-to-one correspondence between F & O
  - Customary metrics can’t necessarily be applied
- Construct MVV so single-valued F & O a special case
  - Some assurance that methods are comparable
- Applicable for continuous or categorical variables
  - Will demonstrate development for continuous variables
  - Analogous for categorical variables
- For example, consider maximum temperature…

The Case for MVV (4)

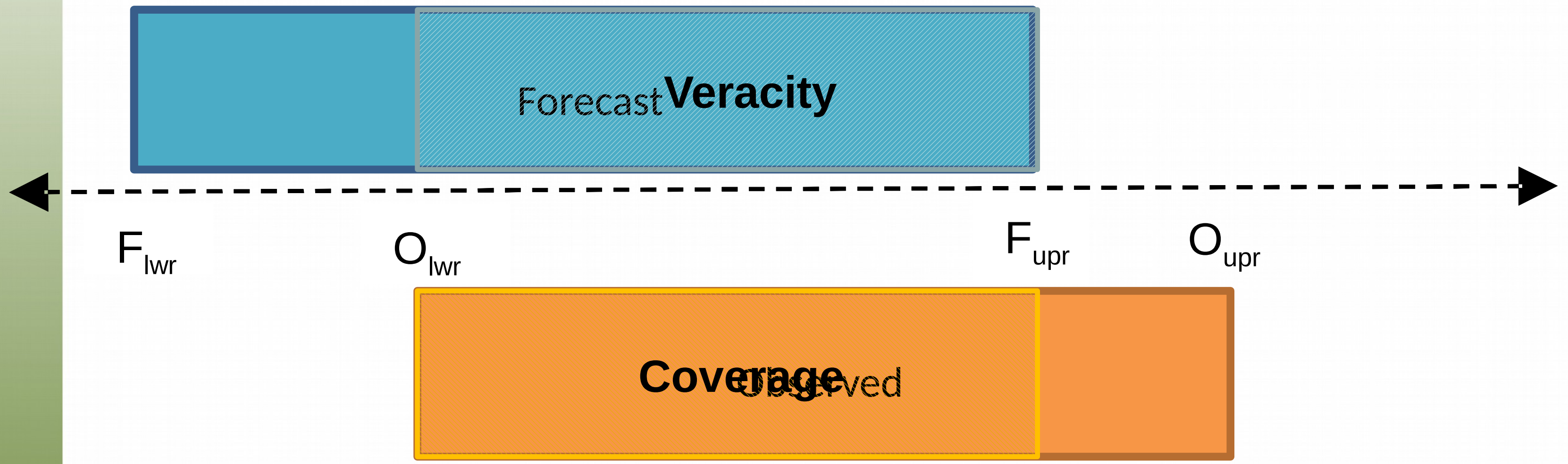
Forecast max temperature range: $[F_{lwr}, F_{upr}]$



Observed max temperature range: $[O_{lwr}, O_{upr}]$

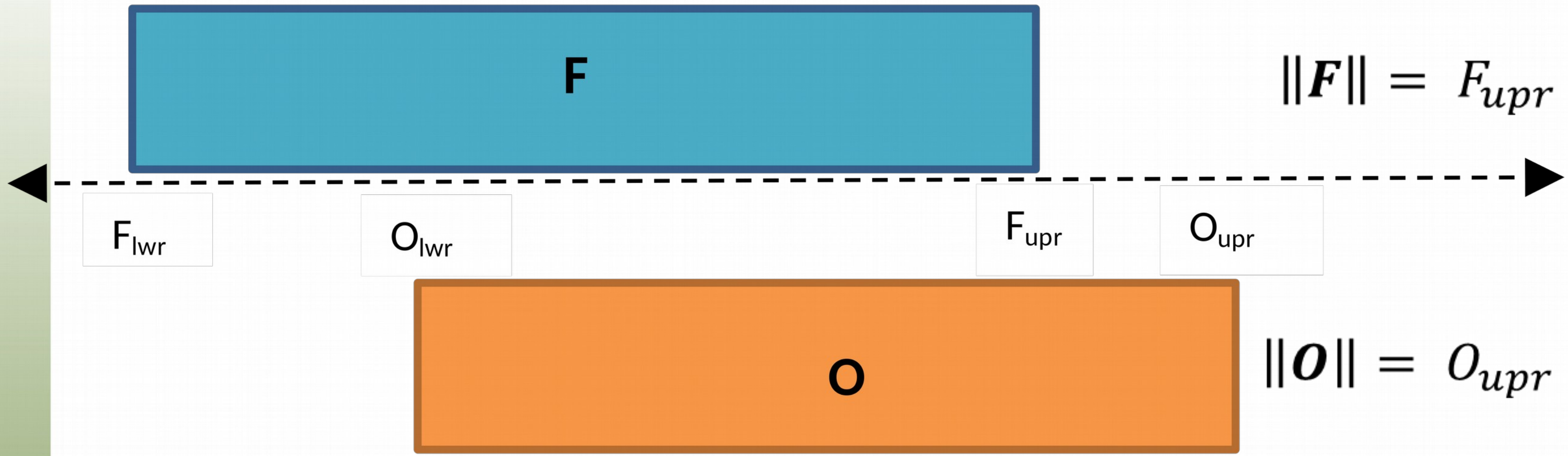
Dimensionless Metrics

Veracity: Fraction of forecast range that was observed



Coverage: Fraction of observed range that was forecast

Dimensionless Metrics (2)



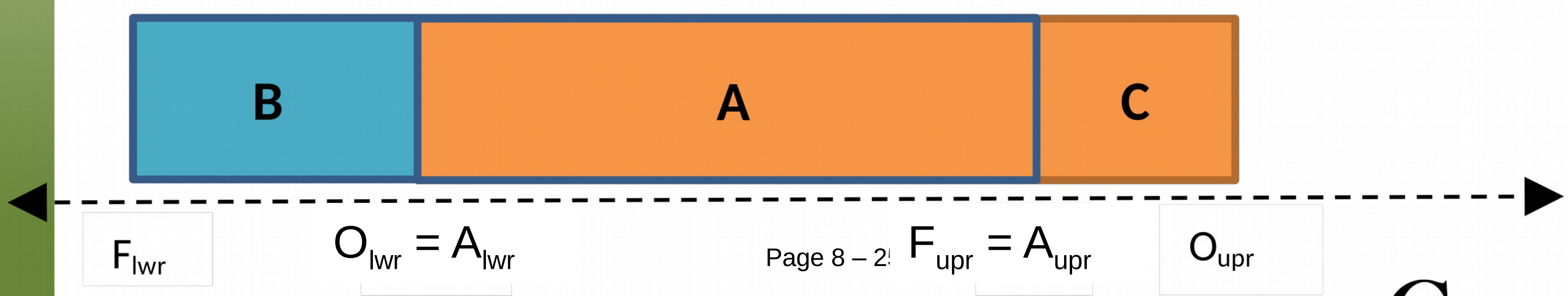
$$||\boldsymbol{F}|| = F_{upr} - F_{lwr}$$

$$||\boldsymbol{O}|| = O_{upr} - O_{lwr}$$

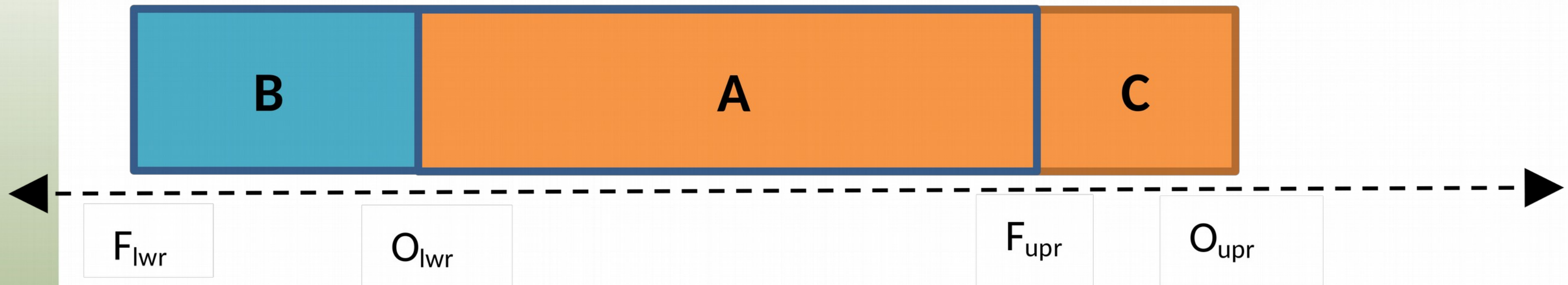
$$A_{lwr} = \max(F_{lwr}, O_{lwr})$$

$$A_{upr} = \min(F_{upr}, O_{upr})$$

$$||\boldsymbol{A}|| = \begin{cases} A_{upr} - A_{lwr} & if \;\; A_{upr} \geq A_{lwr} \\ 0 & if \;\; A_{upr} < A_{lwr} \end{cases}$$



Dimensionless Metrics (3)



$$\|B\| = \|F\| - \|A\| \qquad \|C\| = \|O\| - \|A\|$$

$$Veracity: \frac{\|A\|}{\|F\|} = \frac{\|A\|}{\|A\|+\|B\|} \qquad Coverage: \frac{\|A\|}{\|O\|} = \frac{\|A\|}{\|A\|+\|C\|}$$

- Special cases: $\|F\| = 0$ and/or $\|O\| = 0$
  - Single-valued forecast and/or observation
  - Veracity and/or coverage undefined ($(0 \div 0)$)
  - Example: precipitation amounts when none are forecast/observed

Dimensionless Metrics (4)

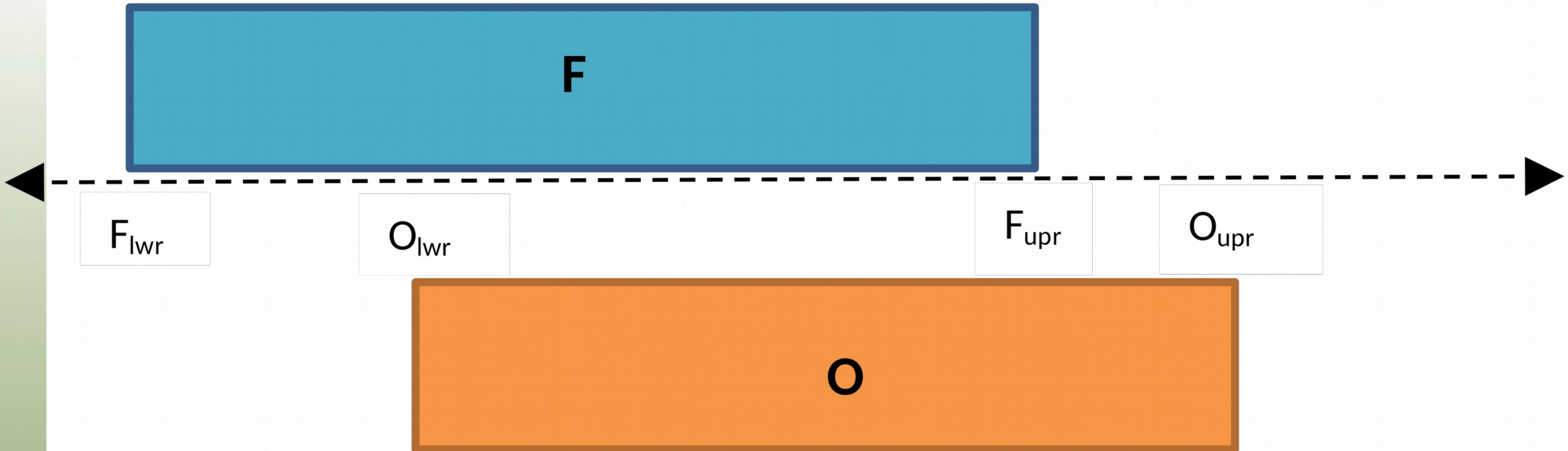
- Define: $\boldsymbol{F} = \hat{\boldsymbol{x}}$ and/or $\boldsymbol{O} = \boldsymbol{x}$
- $\|\boldsymbol{F}\| = 0$ and $\|\boldsymbol{O}\| \neq 0$
  - *Veracity*: $\begin{cases} 1\ if\ \hat{x} \in \boldsymbol{O} \\ 0\ otherwise \end{cases}$; *Coverage*: 0
- $\|\boldsymbol{F}\| \neq 0$ and $\|\boldsymbol{O}\| = 0$
  - *Veracity*: 0 ; *Coverage*: $\begin{cases} 1\ if\ x \in \boldsymbol{F} \\ 0\ otherwise \end{cases}$
- $\|\boldsymbol{F}\| = 0$ and $\|\boldsymbol{O}\| = 0$
  - *Veracity*: $\begin{cases} 1\ if\ \hat{x} = x \\ 0\ otherwise \end{cases}$
  - *Coverage*: $\begin{cases} 1\ if\ \hat{x} = x \\ 0\ otherwise \end{cases}$

Dimensionless Metrics (5)

- Consider veracity and coverage together
  - Can construct measure analogous to Threat Score (CSI):
  - $CSI = \frac{||\boldsymbol{A}||}{||\boldsymbol{A}|| + ||\boldsymbol{B}|| + ||\boldsymbol{C}||}$
  - Similar special case when $||\boldsymbol{F}|| = 0$ and $||\boldsymbol{O}|| = 0$
- Can extend MVV to include correct negatives (missing D)
  - Choice of “D” provides insight into design of forecast system
  - Will not explore further for this example in the interest of time
- Generalization of classical 2x2 contingency table
  - Cell values for individual F/O pairs other than 0 or 1
- Derivation for a single F/O pair
  - Aggregation by “dividing the sums”

- Consider veracity and coverage together
  - Can construct measure analogous to Threat Score (CSI):
  - CSI
  - Similar special case when and
- Can extend MVV to include correct negatives (missing D)
  - Choice of “D” provides insight into design of forecast system
  - Will not explore further for this example in the interest of time
- Generalization of classical 2x2 contingency table
  - Cell values for individual F/O pairs other than 0 or 1
- Derivation for a single F/O pair
  - Aggregation by “dividing the sums”

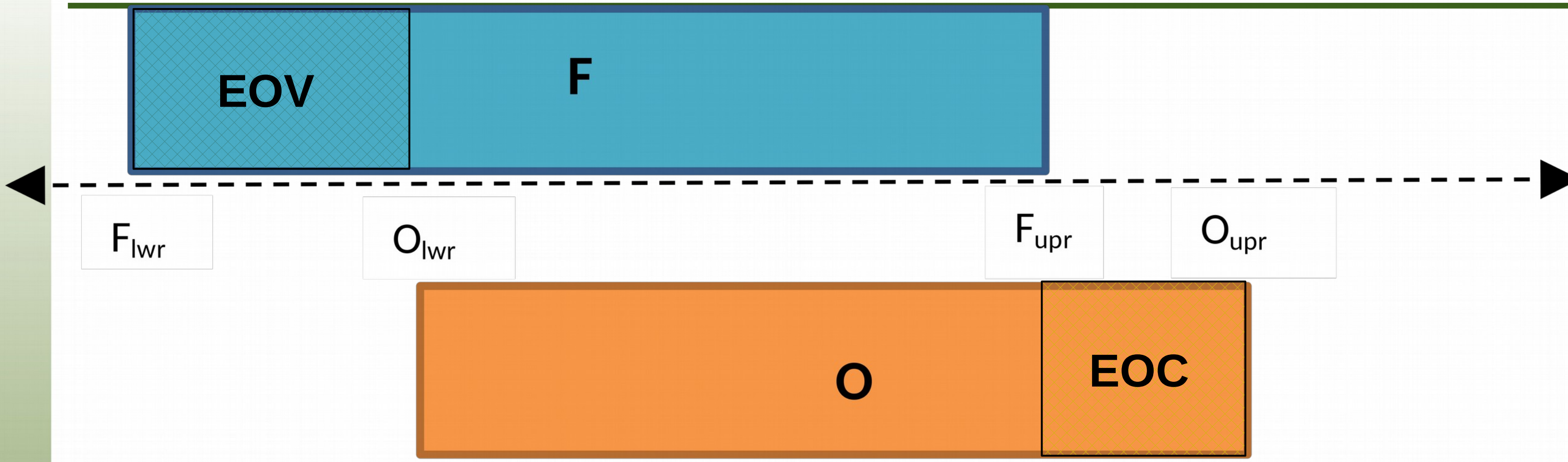
Dimensional metrics



- For continuous variables use customary distance measures
- F and O are defined by end points (Stephenson, 2008)
- Compute standard measures for $(F_{upr}, O_{upr}), (F_{lwr}, O_{lwr})$
  - Bias, MAE, RMSE
- Can compute same measures for means $(\bar{F}, \bar{O})$
- For single-valued system, reduces to measures for (F, O)

- For continuous variables use customary distance measures
- F and O are defined by end points (Stephenson, 2008)
- Compute standard measures for (
  - Bias, MAE, RMSE
- Can compute same measures for means
- For single-valued system, reduces to measures for

Dimensional metrics (2)



- Total absolute error: $|F_{upr} - O_{upr}| + |F_{lwr} - O_{lwr}|$
- Can partition, errors analogous to dimensionless metrics:
  - "Error of veracity": $EOV = |F_{upr} - A_{upr}| + |F_{lwr} - A_{lwr}|$
  - "Error of coverage": $EOC = |O_{upr} - A_{upr}| + |O_{lwr} - A_{lwr}|$
- Single-valued: $2|F - O|$ = error w.r.t. F + error w.r.t O

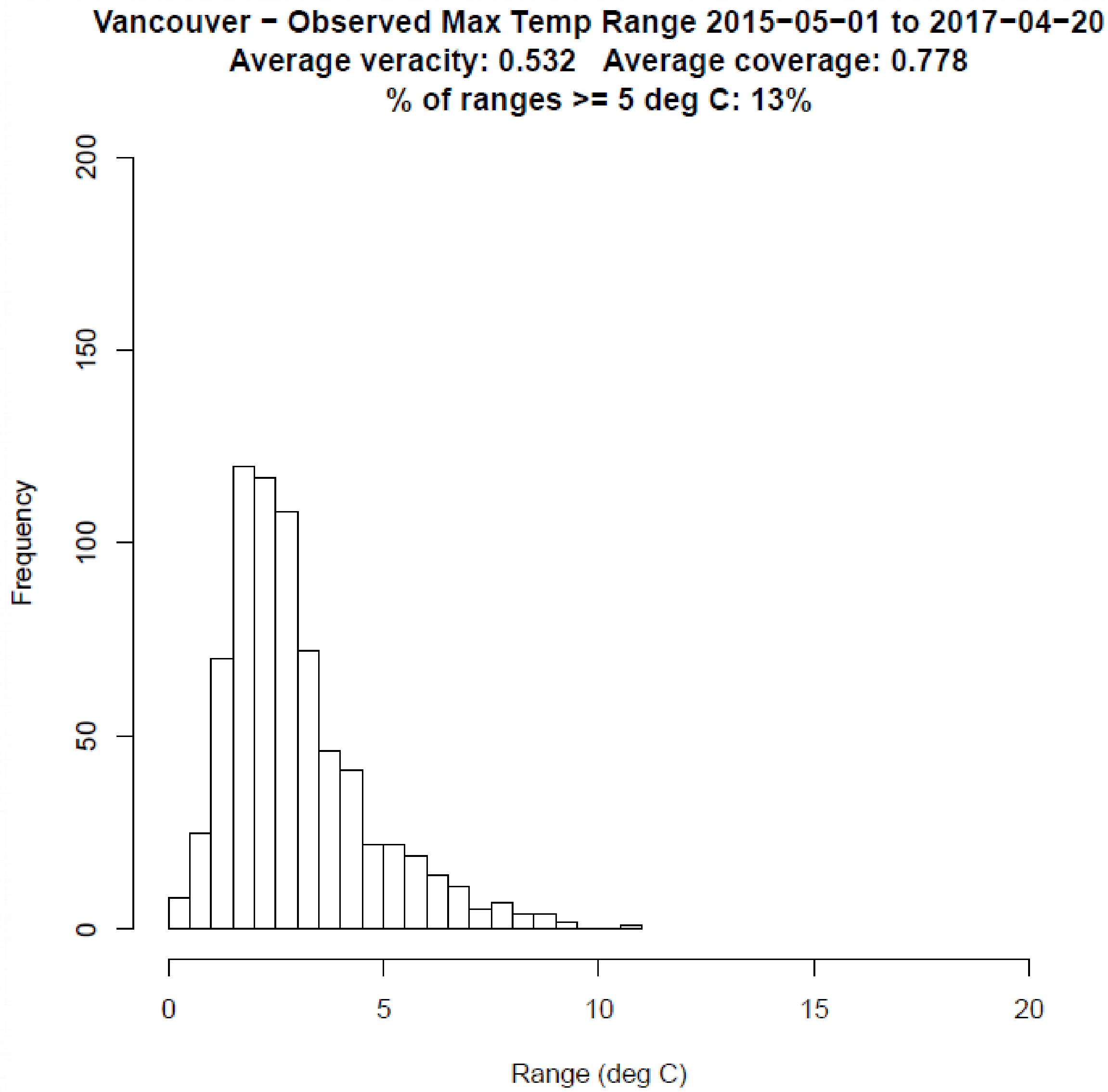
Some applications

- Maximum temperature forecasts
  - Forecasts and observations occupy a range of values
- Marine wind speed forecasts
  - Forecast ranges are not mutually-exclusive
  - Precludes use of categorical verification techniques

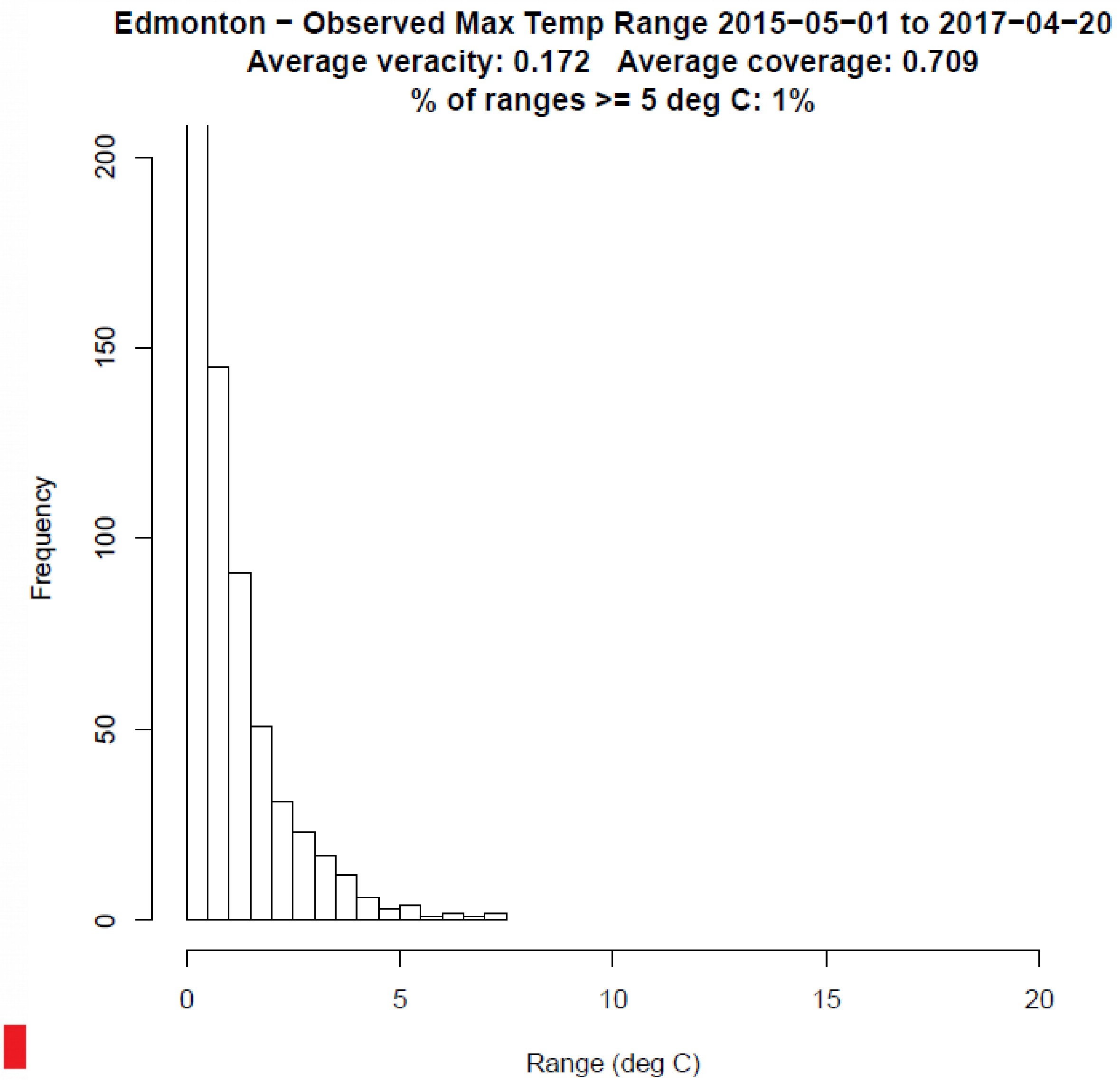
Application: D-1 Max temp forecasts

- Two options for forecast maximum temperature:
  - Single valued: Infer range of ±2°C (in accordance with policy)
    - E.g. “High 15.”
  - “Main condition” with “exception”: use as end points of range
    - E.g. “High 20 except 10 near the coast.”
- Compare D-1 forecasts for 4 Canadian cities for ~2 yrs:
  - Vancouver (West Coast)
  - Edmonton (Prairies, lee of Rocky Mountains)
  - Toronto (Great Lakes)
  - Halifax (East Coast)
- Each has multiple observing stations around region
- Consider frequency distribution of range of observed max imum temperature

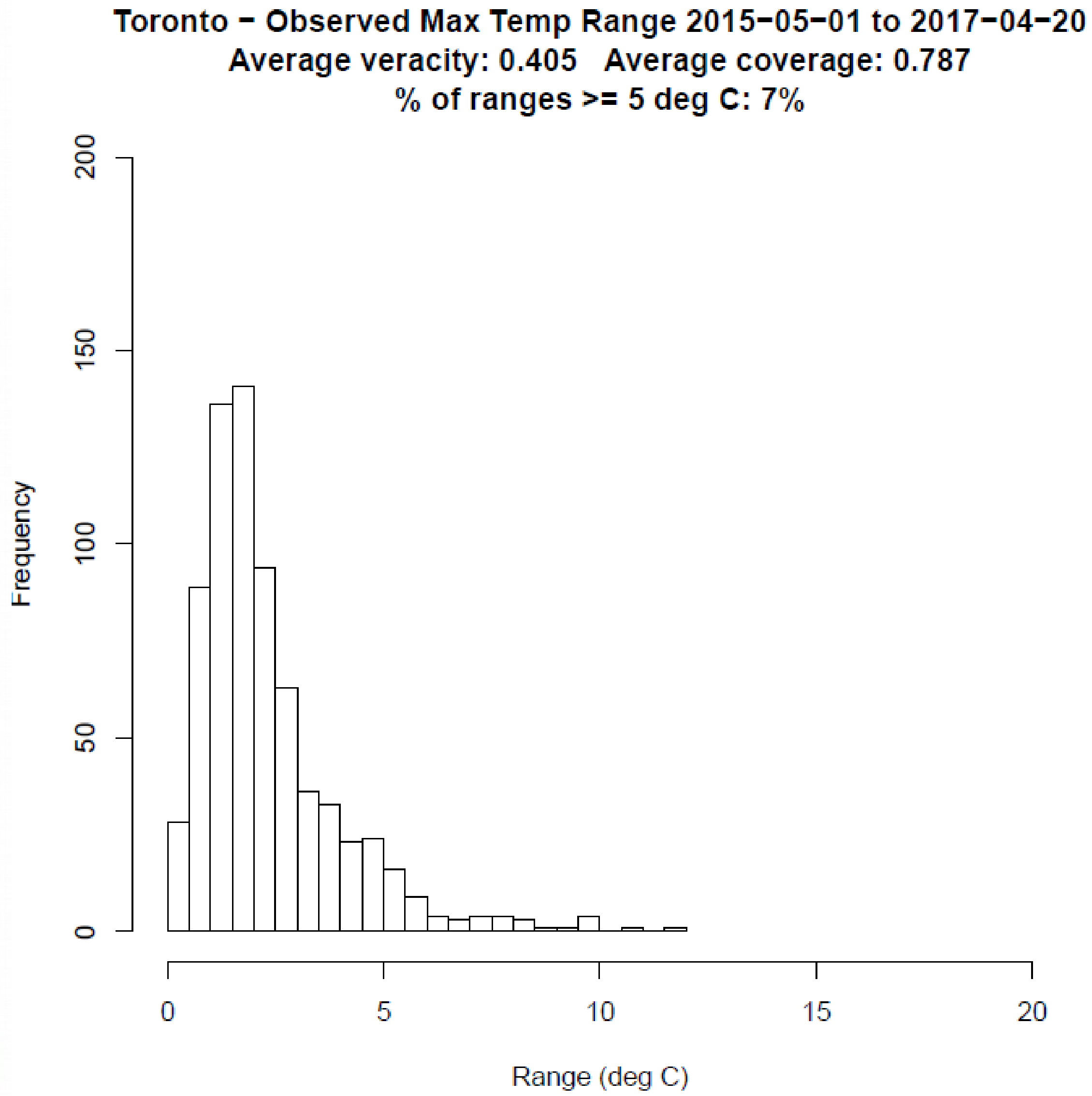
Application: D-1 Max temp forecasts



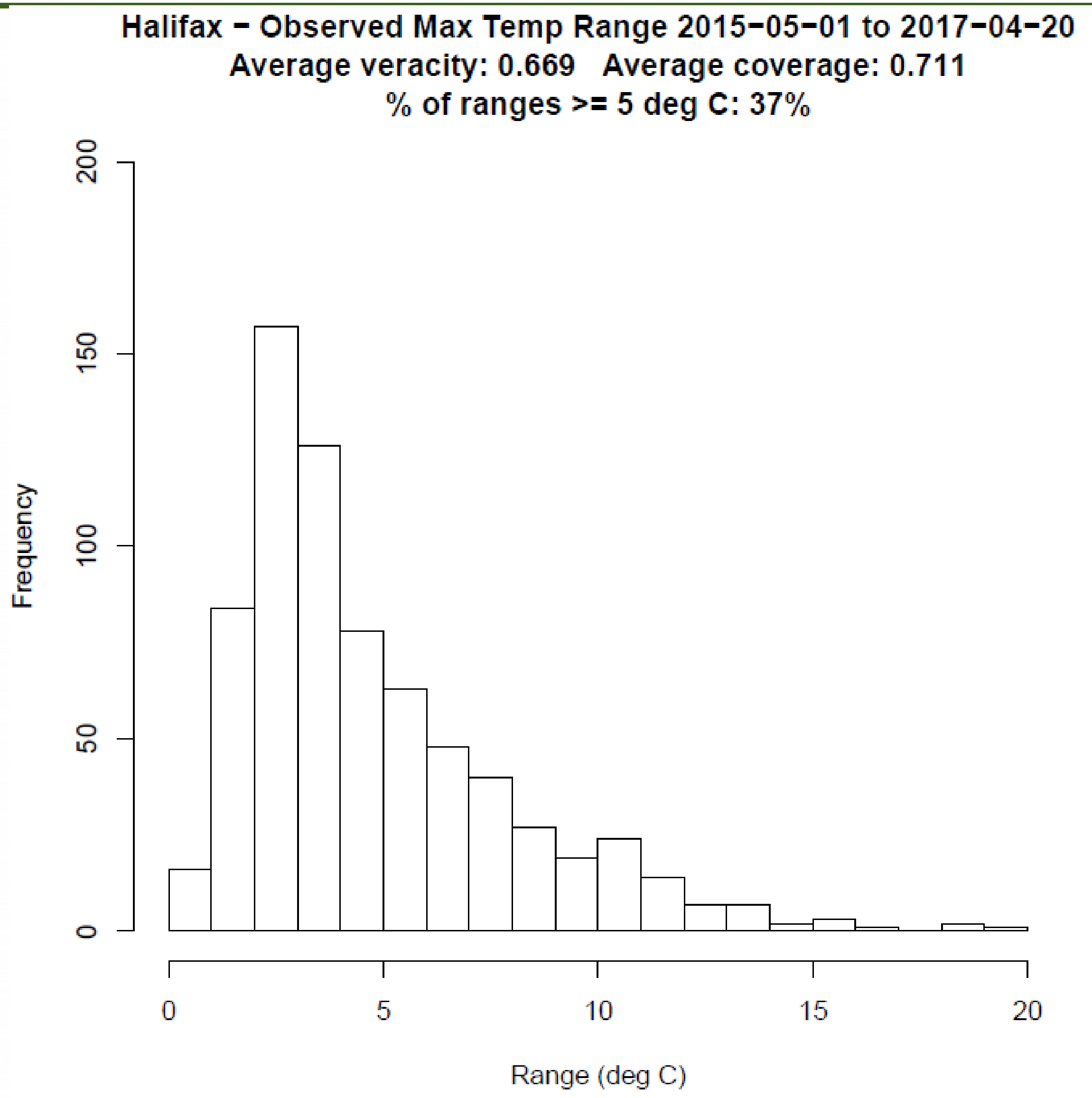
Application: D-1 Max temp forecasts



Application: D-1 Max temp forecasts



Application: D-1 Max temp forecasts



Application: D-1 Max temp forecasts

Forecast Region	Vancouver	Edmonton	Toronto	Halifax
Average Forecast Range (°C)	4.05	4.00	4.24	4.56
Average Observed Range (°C)	3.02	1.00	2.38	4.80
Aggregate Veracity	0.54	0.17	0.42	0.69
Aggregate Coverage	0.72	0.68	0.73	0.65
Aggregate “CSI”	0.44	0.16	0.36	0.51
Bias of Lower Bound (°C)	-1.00	-1.58	-0.37	0.32
Bias of Midpoint of Range (°C)	-0.48	-0.09	0.55	0.18
Bias of Upper Bound (°C)	0.03	1.40	1.48	0.03
Average Error of Veracity (°C)	1.89	3.45	2.53	1.42
Average Error of Coverage (°C)	0.86	0.47	0.68	1.71

Application: 24-30h wind forecasts

- Two options for forecast marine winds:
  - Single valued: Infer range of ±5 knots (in accordance with polic y)
    - E.g. “Wind 15 knots.”
  - Explicit range: use as end points of range
    - E.g. “Wind 15 to 20 knots.”
- Consider marine forecast for 2 regions for ~3 months:
  - Georgian Bay (Great Lakes)
  - Banquereau (Western North Atlantic)
- Observations from single buoy over 6-hour period

Application: 24-30h wind forecasts

Forecast Region	Georgian Bay	Banquereau
Average Forecast Range (knots)	12.3	11.6
Average Observed Range (knots)	5.3	6.0
Aggregate Veracity	0.35	0.28
Aggregate Coverage	0.79	0.55
Aggregate "CSI"	0.32	0.23
Bias of Lower Bound (knots)	-1.04	2.03
Bias of Midpoint of Range (knots)	2.32	4.87
Bias of Upper Bound (knots)	5.68	7.71
Average Error of Veracity (knots)	7.84	9.13
Average Error of Coverage (knots)	1.12	3.45

Future work

- Apply method to other forecast fields
  - Public forecast: wind, precipitation amount, precipitation type
  - Marine forecast: wind direction, wave height, air temperature
  - Aviation forecast: precipitation type, categories
- Verification of existing forecasts using big data
  - E.g. satellite observations, synthetic observations
- Gridded forecast verification
  - Define forecast range for grid boxes from grid points
- Extensions of basic method
  - Two dimensions, e.g. verification of vector wind
  - Weighting of forecasts and/or observations