Best-fit Statistical Distribution and Regionalization Techniques for Flood Frequency Analysis in Canada

Ziyang Zhang[a], Donald H. Burn[b], Tricia A. Stadnyk[c], Peter F. Rasmussen[d] (Deceased)

[a]M.Sc. Student, University of Manitoba; [b]Professor, University of Waterloo; [c]Associate Professor, University of Manitoba; [d]Professor, University of Manitoba

umzha494@myumanitoba.ca, dburn@uwaterloo.ca, tricia.stadnyk@umanitoba.ca

UNIVERSITY OF MANITOBA

NSERC Canadian FloodNet
Annual General Meeting, 2017
Theme 1, Project 1-6

FloodNet NSERC

1. Introduction

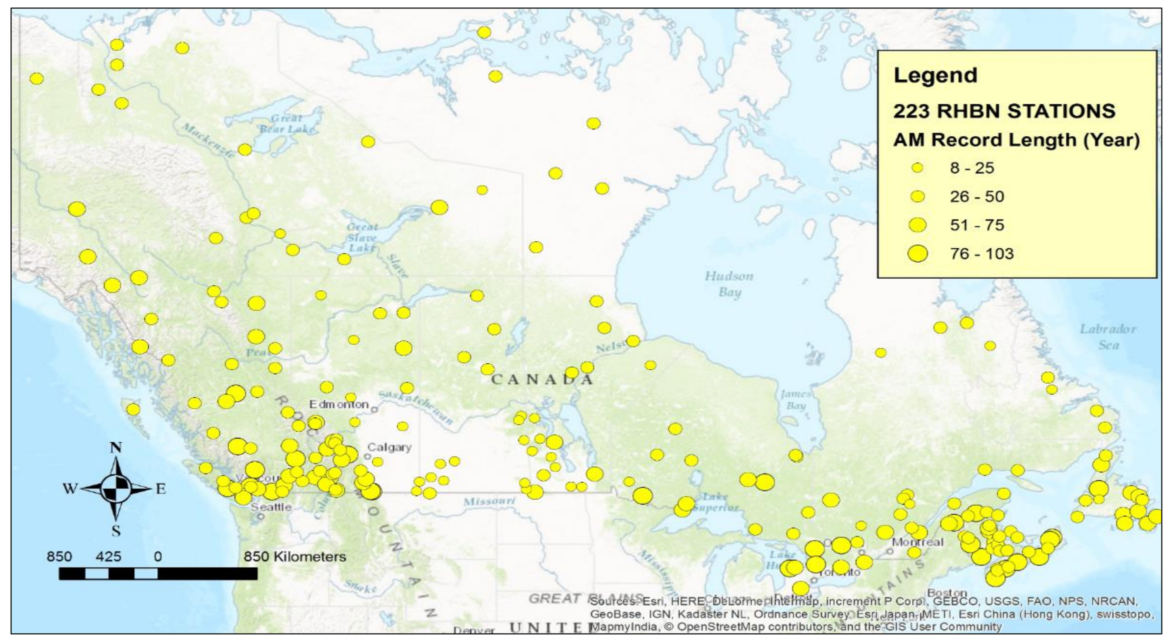
Fitting annual maximum flood series to a statistical distribution in the form of flood frequency analysis is widely used to estimate design flood for large return period. The key to the success of this approach is dependent on: i) the type of distribution being used, and ii) the lengths of the flood record. If the length of a flood sample is relatively too short, regional flood frequency analysis could carry out in the form of index flood method with region of influence approach (Burn, 1990).

In Canada, there is no national flood estimation guideline to unify the procedures of flood quantile estimation, and thus, no agreement on which statistical distribution should be used and what most effective regionalization technique should be applied.

This project, under Theme 1-6: development of new flood estimation manual for Canada, aims to:
Task i: Find the best-fit distribution for at-site flood frequency analysis in Canada;
Task ii: Investigate techniques to delineate homogeneous pooling groups for regional flood frequency analysis in Canadian climate.

2. Data

Annual maximum (AM) flow data obtained from Environment and Climate Change Canada (ECCC) Reference Hydrometric Basin Network (RHBN) stations



223 stations for task i
- Mean 47.5 year
- Max 103 year
- Min 8 year

186 stations for task ii

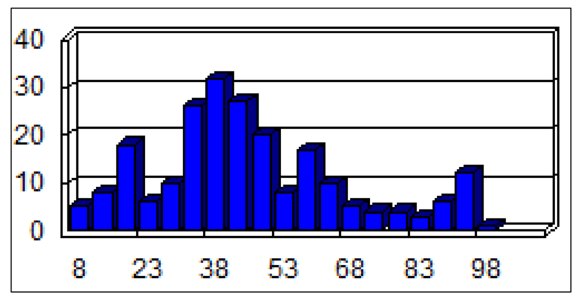
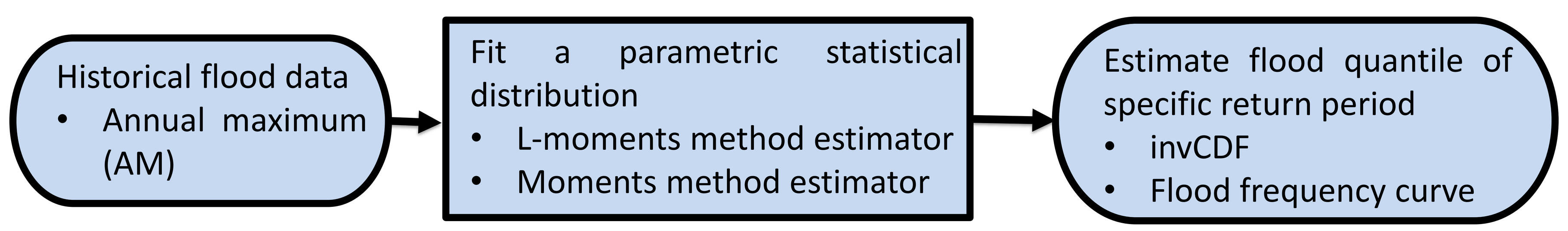


Figure 1: Locations of 223 hydrometric stations

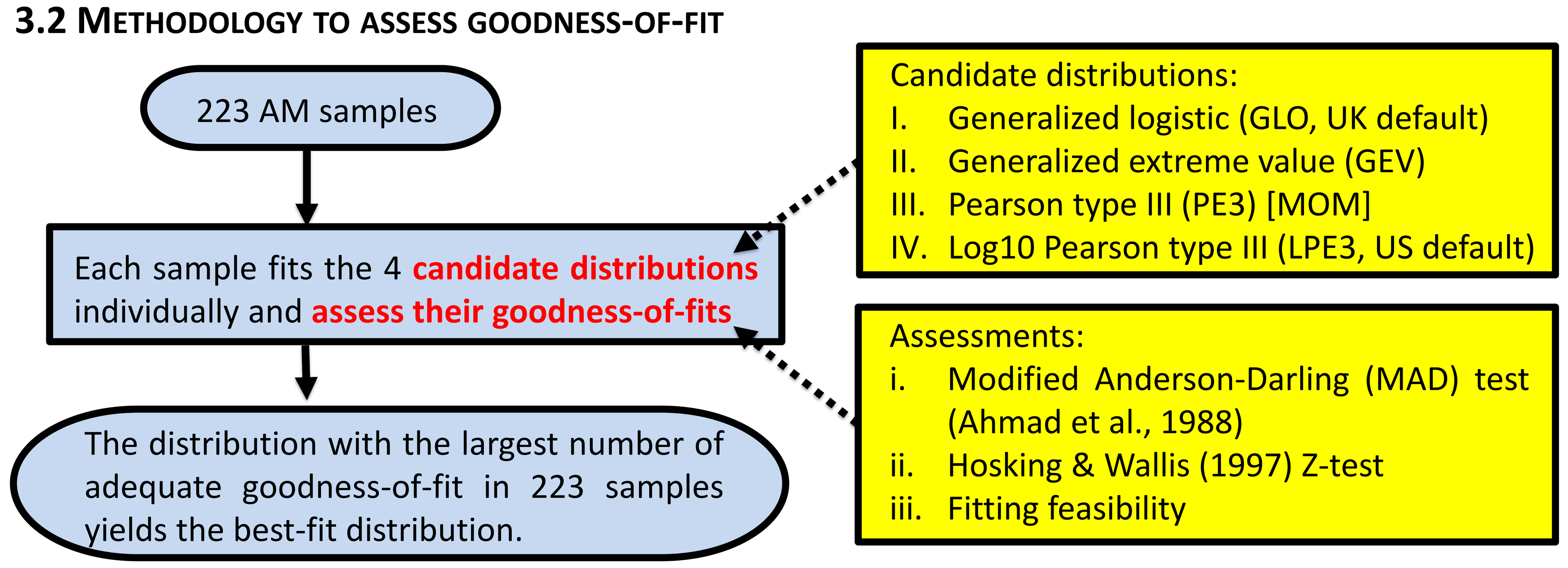
Figure 2: Frequency distribution of 223 RHBNs record length

3. Task I – find the best fit distribution

3.1 Background: procedures for at-site flood frequency analysis



3.2 Methodology to assess goodness-of-fit



3.4 Modified Anderson-Darling goodness-of-fit test

- It measures the difference between the empirical cumulative distribution function (CDF) and the estimated cumulative distribution function.
- Empirical CDF is a step function that jump up 1/n at each of the n observational values.
- It is a modified version of Anderson-Darling test that emphasizes on measuring the discrepancies on the upper tail of the two CDFs which corresponding to flood quantile of large return period.
- The output is a test statistics value and is compared with critical values to decide whether the estimated distribution is "REJECT" or "DO NOT REJECT".

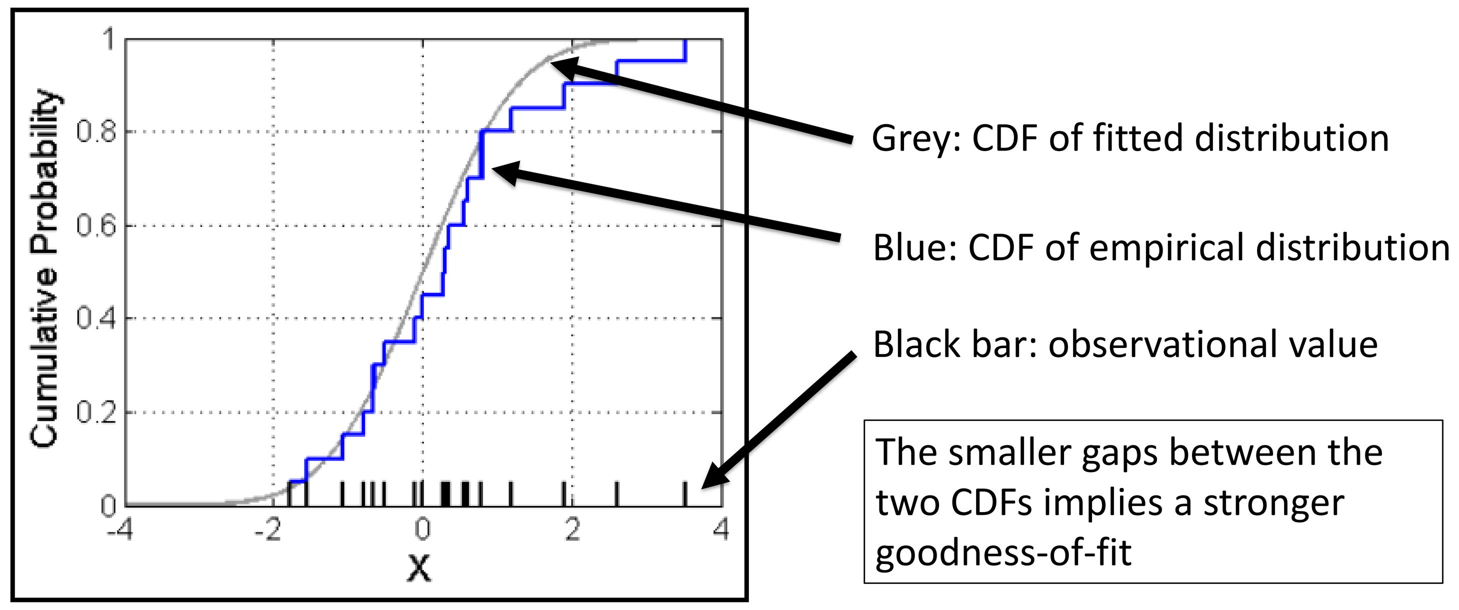


Figure 3: Graphical illustration of the theory of MAD test

3.5 Hosking & Wallis Z-test

- Measure distance between sample dot and distribution curve in the L-Skewness to L-Kurtosis 2-D spaces.
- Closer distance between the sample estimated dot and the distribution curve implies a stronger goodness-of-fit.
- The output is Z value. -1.64 < Z < 1.64 suggests adequate goodness-of-fit.

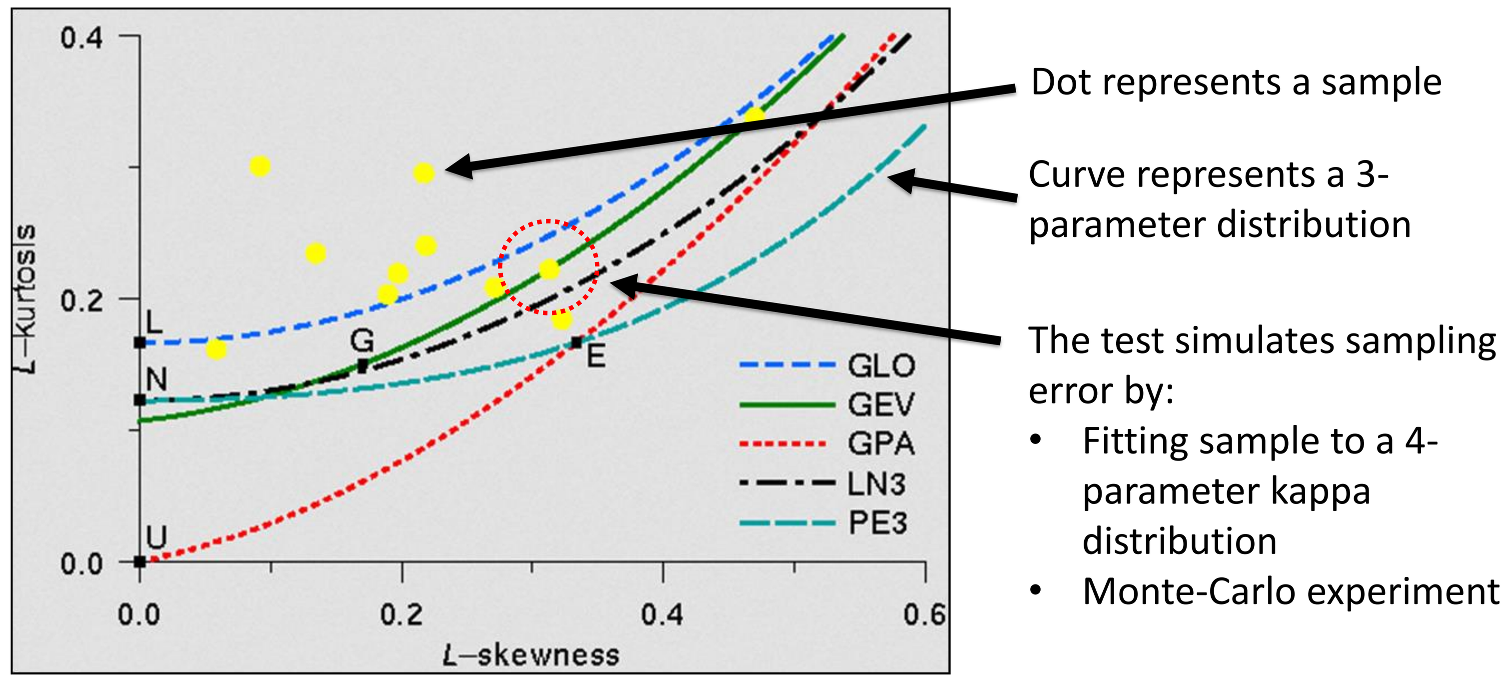


Figure 4: Graphical illustration of the theory of Hosking & Wallis Z test

3.6 Result & conclusion of task I

- GEV has better performance compared to other candidates.
- No geographical pattern was found.
- General good agreement were found between MAD and Z tests.

Table 1: Number of stations which their flood data are feasible to fit and adequate to fit

	Annual Maximum 223 RHBN Stations in total				Instantaneous Peak 204 RHBN Stations in total			
Assessment	**GLO**	**GEV**	**PE3**	**Log10 PE3**	**GLO**	**GEV**	**PE3**	**Log10 PE3**
Feasible to fit distribution	220	223	176	211	199	202	195	186
Adequate fitting by Z-Test at 0.1	168	194	197	184	153	175	150	161
Fitting "Do not reject" by MAD test at 0.1	180	192	149	183	171	177	133	172

3.3 Fitting feasibility limited by distribution boundary

Example of GLO:

CDF: $F(x) = 1 / (1 + e^{-y})$

where $y = -k^{-1} * \log \boxed{\{1 - k(x - \varepsilon)/\alpha\}}$ if $k \neq 0$

$y = (x - \varepsilon)/\alpha$ if $k = 0$

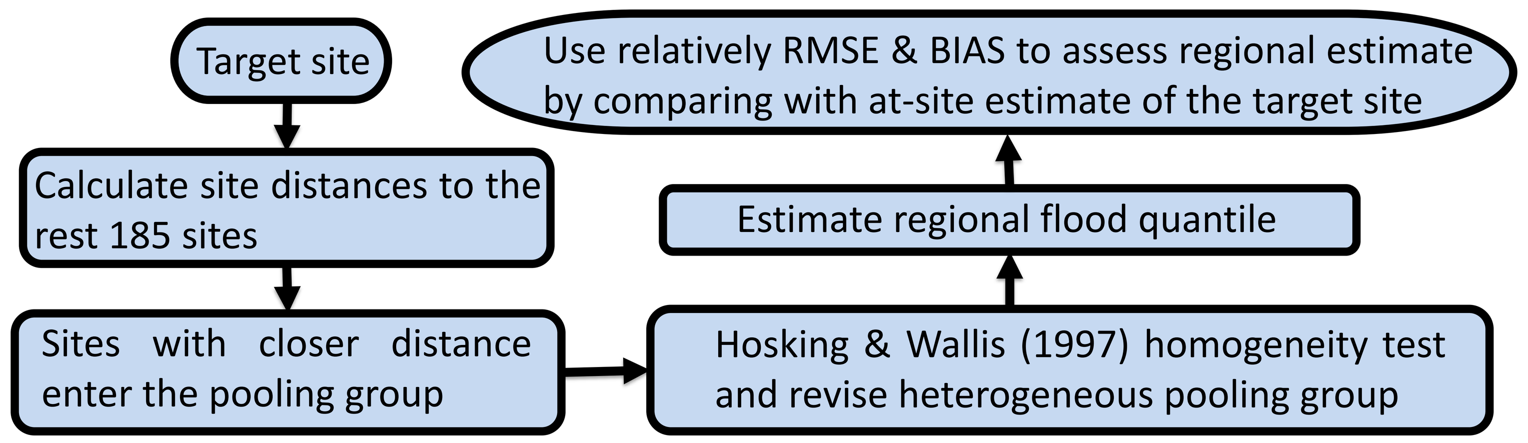
Needs to be greater than 0

k: shape parameter, α: scale parameter, & ε: location parameter

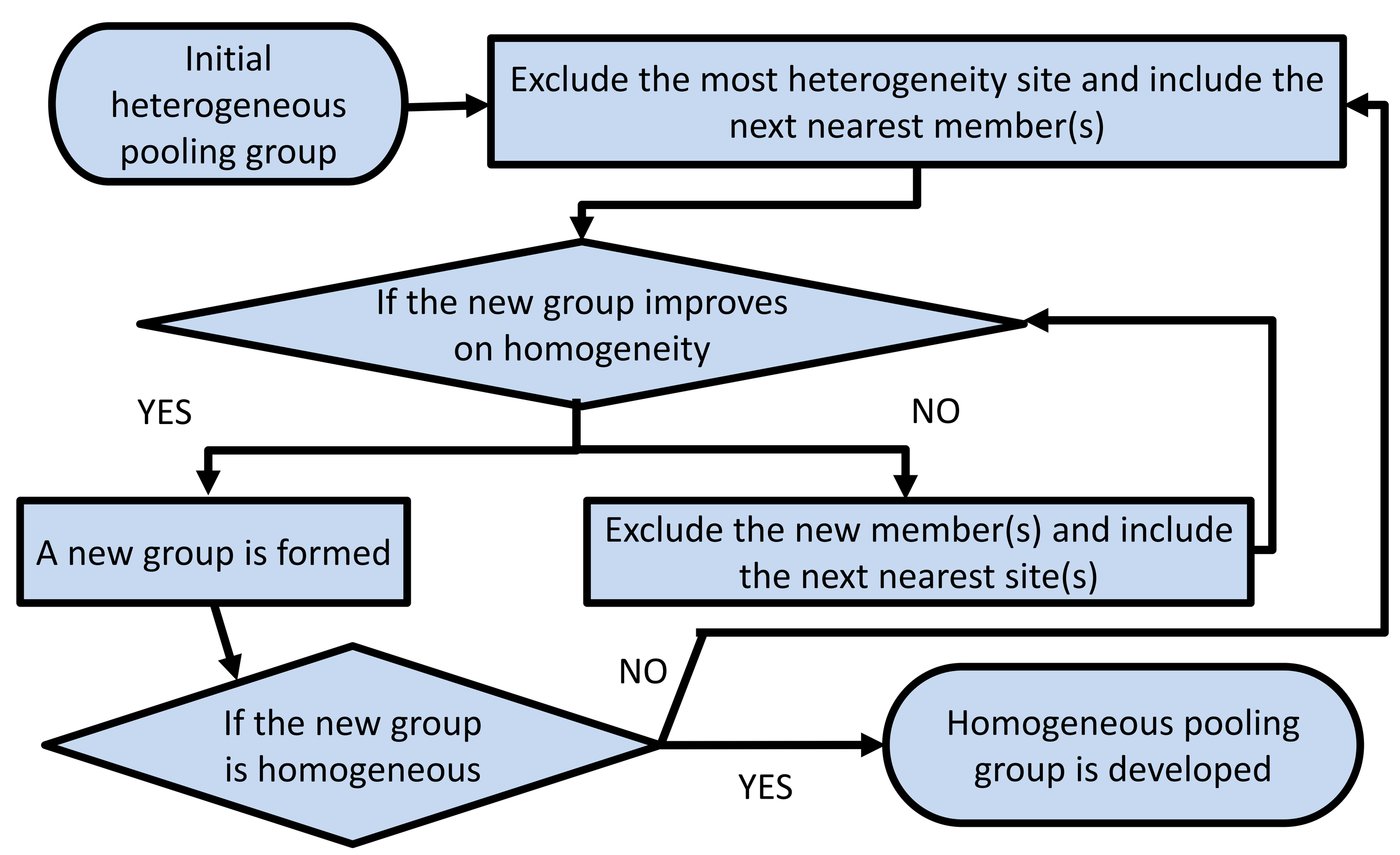
4. Task II delineate homogeneous pooling group for regional flood frequency analysis

- Develop homogenous pooling group for each target site (186 target sites in total).
- Homogeneous pooling group: flood samples in the group have an identical distribution and differ only in terms of a scale parameter.
- Use and compare 6 different approach to measure the distance of similarity between stations.
  i. Flood seasonality (Burn, 1997)
  ii. Physiographic characteristics – selected by stepwise regression model: BASIN_AREA, WatBody_MeanSize, T_AVG_BASIN, ELEV_STD, PPTAVG_BASIN
  iii. Proximity (Station LAT, LON)
  iv. Monthly average precipitation (based on correlation coefficient)
  v. Monthly average temperature (based on correlation coefficient)
  vi. iv + v

4.2 Procedures to develop pooling group for each target site



4.3 Procedures to revise heterogeneous pooling group



4.4 Result and conclusion of task II

- The key that effectively develops homogeneous pooling groups is less dependent on the characteristics used to measure flood similarity, but relies more on the regionalization process; particularly the procedure to revise a heterogeneous pooling group.
- All regionalization approaches yield similar performance in terms of relatively RMSE & BIAS measures.
- 9 prairie stations could not form homogeneous pooling groups.

Table 2: Percentage of homogeneous pooling group formed by each techniques in a total of 186 stations

Techniques	Flood seasonality	Physiographic characteristics	Proximity	Monthly avg. prec.	Monthly avg. temp.	Monthly avg. prec. + temp.
H<1	94.1%	91.9%	87.1%	89.8%	87.6%	90.9%

Table 3: Relative RMSE & BIAS measures between regional & at-site estimate for 19 stations (AMs>90 year)

	RMSE						BIAS [%]					
T	Flood seasonality	Physiographic characteristics	Proximity	Monthly avg. prec.	Monthly avg. temp.	Monthly avg. prec. + temp.	Flood seasonality	Physiographic characteristics	Proximity	Monthly avg. prec.	Monthly avg. temp.	Monthly avg. prec. + temp.
10	0.041	0.039	0.04	0.041	0.039	0.04	1.73	0.84	1.9	2.22	1.58	1.95
25	0.077	0.07	0.063	0.071	0.068	0.075	2.9	1.27	2.33	3.36	1.79	3.08
50	0.11	0.099	0.085	0.101	0.095	0.107	3.76	1.6	2.52	4.16	1.86	3.91
100	0.147	0.128	0.109	0.133	0.123	0.141	4.64	1.96	2.66	4.95	1.91	4.76
250	0.199	0.169	0.142	0.179	0.162	0.189	5.83	2.51	2.82	6.04	1.99	5.96
500	0.242	0.199	0.168	0.215	0.192	0.227	6.78	3.02	2.96	6.92	2.1	6.97

Acknowledgments

This project is supported by the Natural Sciences and Engineering Research Council of Canada under FloodNet Project. Basin physiographic characteristics were provided by Erika Klyszejko (Environment and Climate Change Canada).

References

I. Ahmad, M. I., Sinclair, C. D., & Spurr, B. D. (1988, August). Assessment of Flood Frequency Models Using Empirical Distribution Function Statistics. *Water Resources Research, 24*(8), 1323-1328.
II. Burn, D. H. (1990, October). Evaluation of Regional Flood Frequency Analysis with a Region of Influence Approach. *Water Resources Research, Vol. 26*, 2257-2265.
III. Burn, D. H. (1997). Catchment similarity for regional flood frequency analysis using seasonality measures. *Journal of Hydrology*, *202*(1), 212-230.
IV. Hosking, J. R., & Wallis, J. R. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments.* Cambridge, UK: Cambridge University Press.
V. Robson, A. J., & Reed, D. W. (1999). *Statistical procedures for flood frequency estimation. Volume 3 of the Flood Estimation Handbook.* Oxfordshire: Centre for Ecology & Hydrology.
VI. US Water Resources Council (1976). Guidelines for Determining Flood Flow Frequency - Bulletin 17b. Washington, D.C.: Hydrol. Comm.