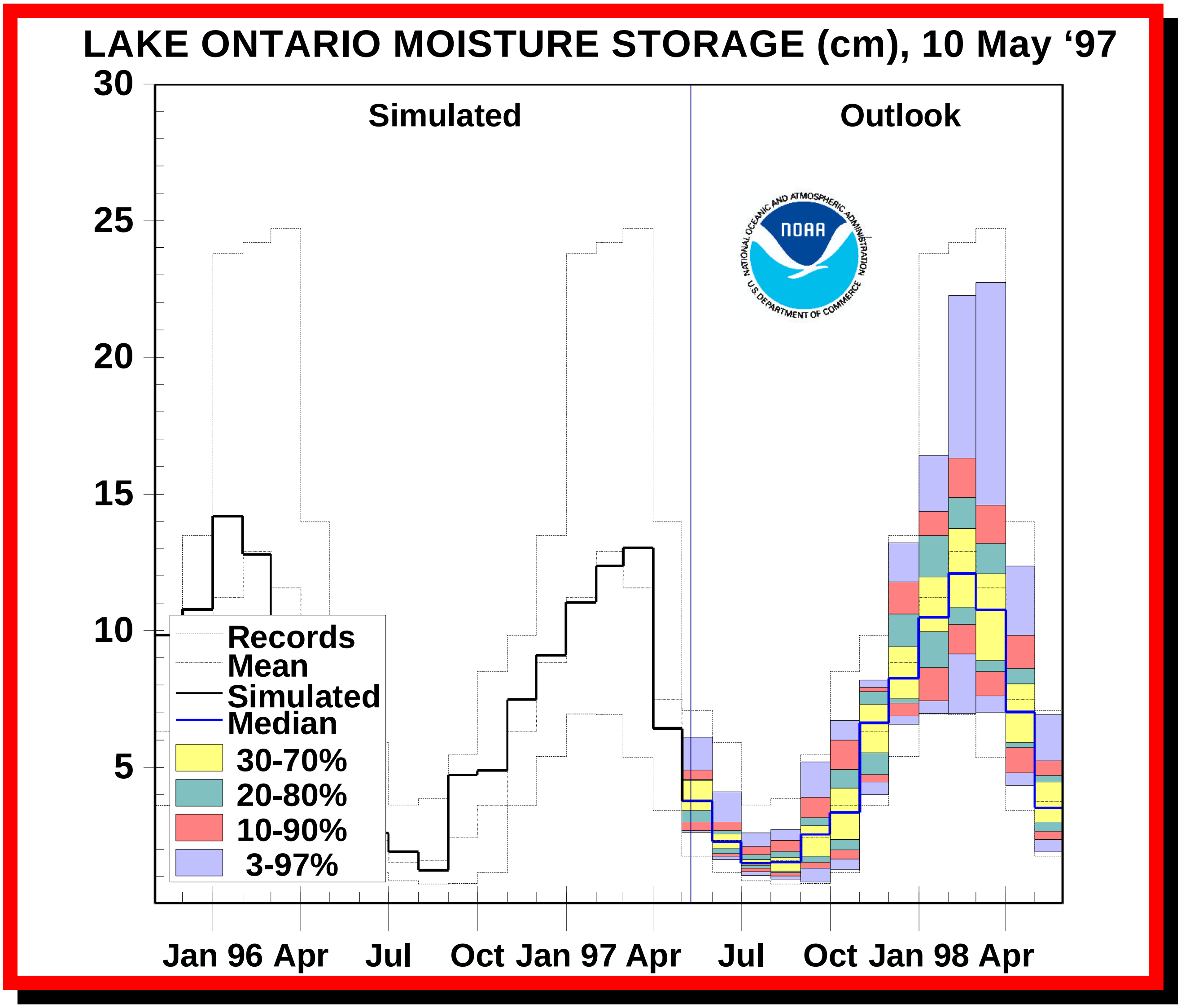
Making Probabilistic Hydrology Outlooks from Probabilistic Meteorology Outlooks



Tutorial Notes

10 July 1997

Thomas E. Croley II
Great Lakes Environmental Research Laboratory
2205 Commonwealth Blvd.
Ann Arbor, Michigan 48105-1593

Making Probabilistic Hydrology Outlooks

from

Probabilistic Meteorology Outlooks

Multiple long-lead probabilistic meteorology outlooks are now available to the water resource engineer or hydrologist. These outlooks are defined over different time periods at different lag times, and they forecast either event probabilities or only most-probable events. An operational hydrology approach is described here for generating probabilistic hydrology outlooks that are compatible with these meteorology outlooks, in order of user priority, yet preserves spatial and temporal relationships observed in past meteorology by using historical data. The approach builds a hypothetical very large structured set of possible future scenarios, with meteorological relative frequencies matching the outlook probabilities, to be treated as a "sample" from which to estimate outlook probabilities and other parameters. The use of this hypothetical set corresponds to the weighted use of a scenario set corresponding to the historical data. Boundary condition equations for the weights are constructed corresponding to forecast event probabilities, and boundary condition inequalities are constructed corresponding to forecast most-probable events. The resulting set of all boundary condition equations is solved for physically-relevant values. The solution may involve an optimization when there are more than one set of weights possible.

These tutorial notes are designed to introduce definitions, available meteorology outlooks, operational hydrology methodology, use of meteorology outlooks as boundary conditions for the determination of weights used in making hydrology outlooks, and solution mechanics for the boundary condition equation set.

Great Lakes Environmental Research Laboratory
National Oceanic and Atmospheric Administration
U.S. Department of Commerce
2205 Commonwealth Blvd.
Ann Arbor, Michigan 48105-1593
USA

Table of Contents

Deterministic Outlooks

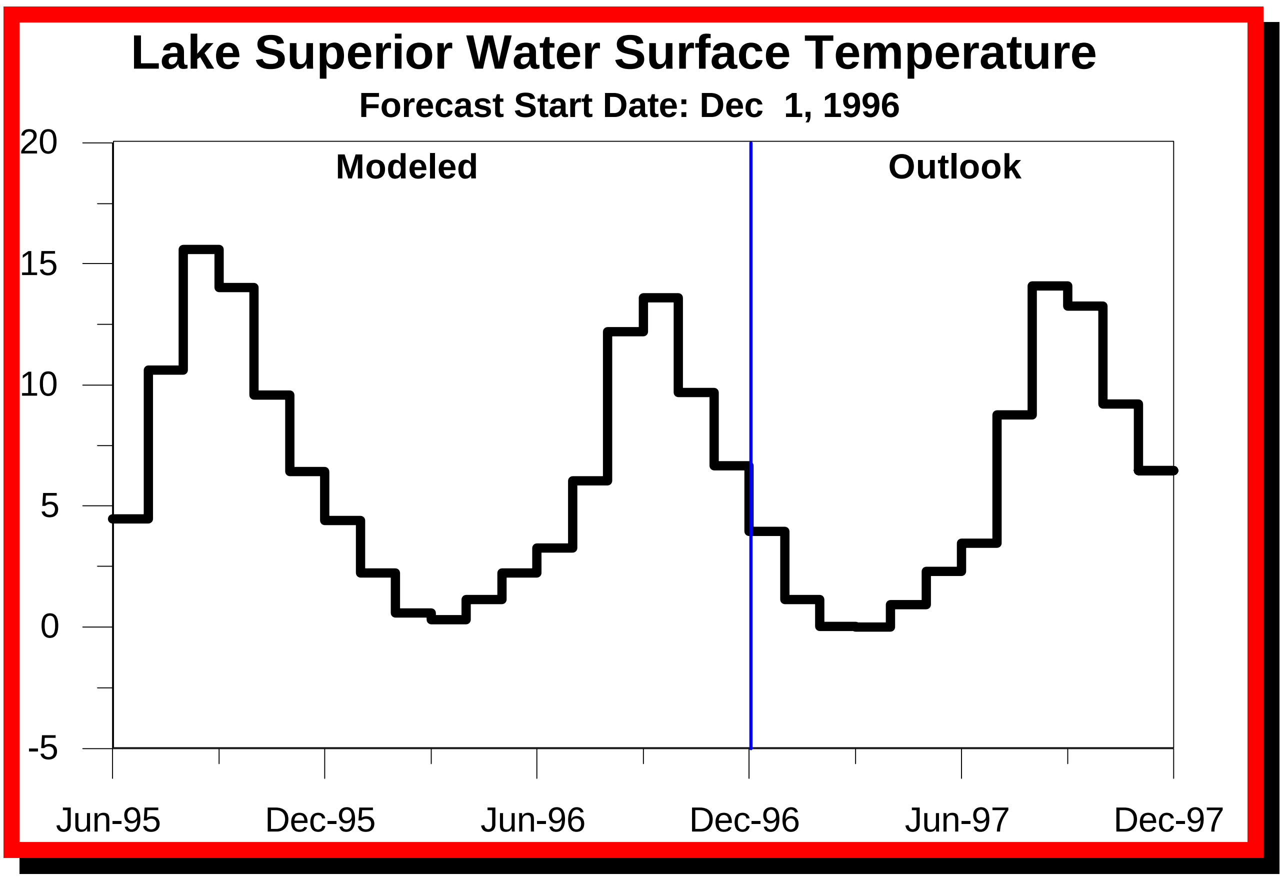
What is a deterministic outlook? Example deterministic meteorology outlooks include the following, made on 1 December for December:

- The temperature in December will be -5C.
- The precipitation in December will be 6 cm.

Example deterministic hydrology outlooks include the following, made on 1 December for December:

- The basin moisture in December will be 21 cm.
- The basin runoff in December will be 3 cm.
- The lake surface temperature in December will be 4C.

An example multiple-valued deterministic hydrology outlook made 1 December is that the water temperature for the next 12 months will be as pictured:



Advantages of deterministic outlooks are that they are easy to understand, easy to make, and easy to assess. Disadvantages are they are difficult to make correctly, they have low information content, and there is a need to select the "best" when multiple outlooks are present.

Probabilistic Outlooks

What is a probabilistic outlook? Example probabilistic meteorology outlooks include the following, made on 1 December for December:

Probability that December temperature exceeds -3C will be 25%;

$$P[Temp_{Dec} > -3C] = 0.25$$

The probability December precipitation is below 5 cm will be 40%;

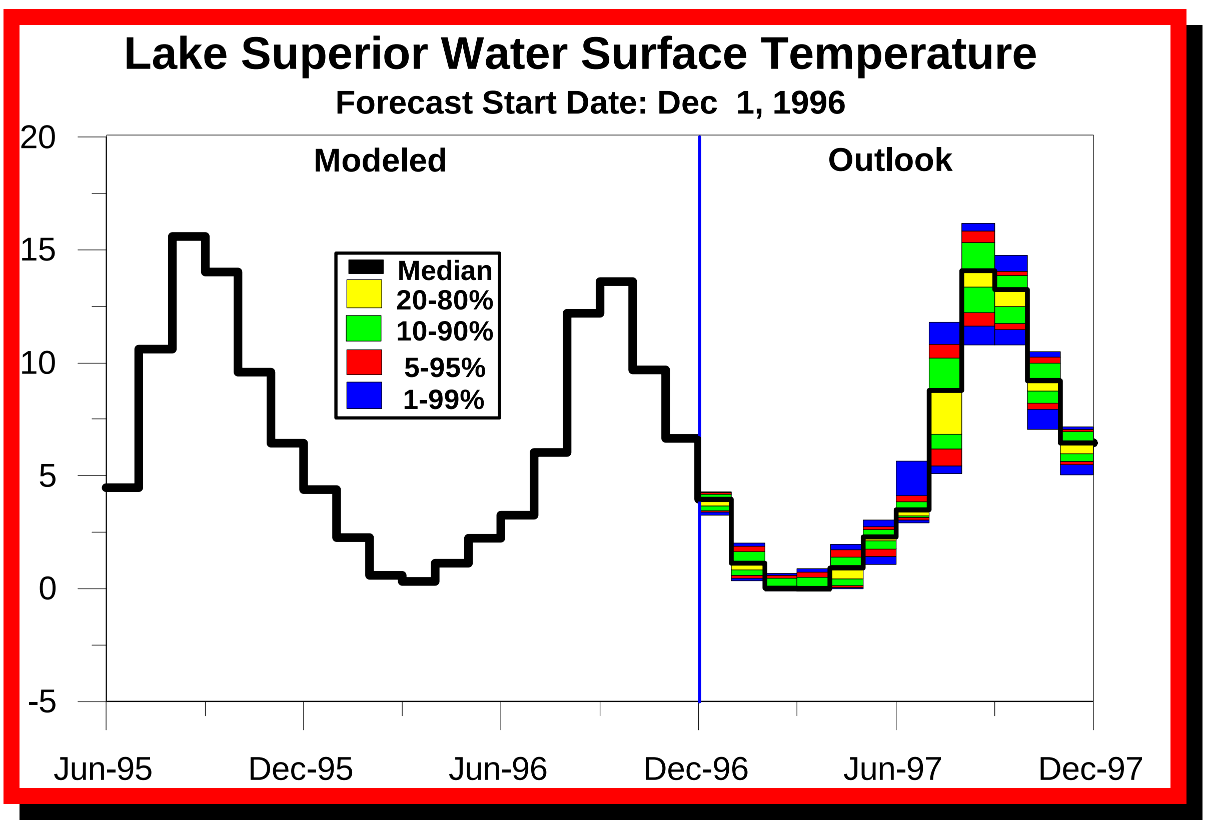
$$P[Prec_{Dec} < 5\ cm] = 0.40$$

Example probabilistic hydrology outlooks include the following made on 1 December for December:

$$P[Runoff_{Dec} < 3.5\ cm] = 0.65$$

$$P[3.4C < WaterTemp_{Dec} < 4.2C] = 0.90$$

An example multiple-valued probabilistic hydrology outlook made 1 December is that water temperature probability distributions for the next 12 months will be as pictured:

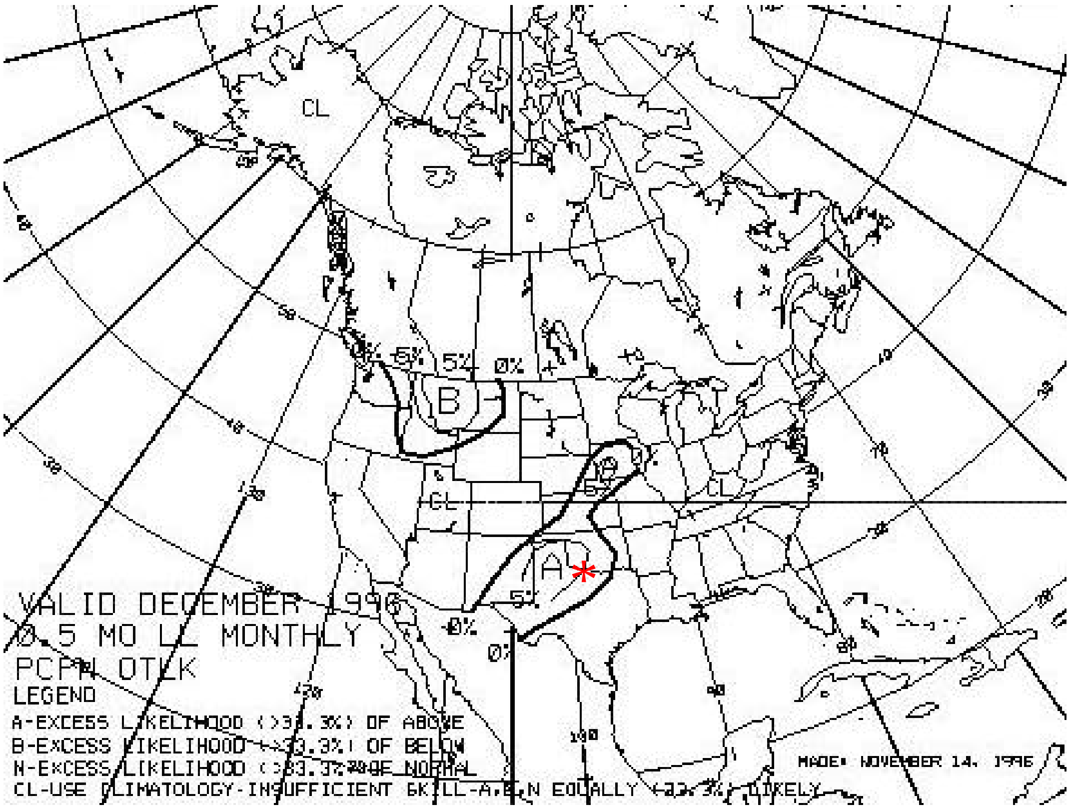


Disadvantages of probabilistic outlooks are that they are difficult to understand, difficult to make, and difficult to assess. Advantages are they have high information content, maximizing the use of available information and providing more information for decision makers. They also obviate the need to select a "best" deterministic outlook.

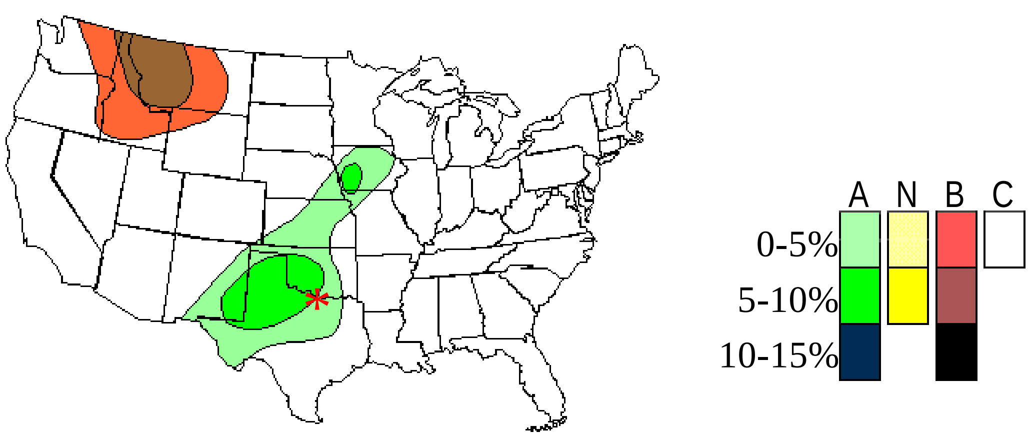
Multiple Probabilistic Meteorology Outlooks

NOAA 1- & 3-Month Climate Outlooks

What are multiple probabilistic meteorology outlooks? They can consist of multiple probabilities within one outlook; for example, the NOAA 1-month climatic outlook for December, made 14 November 1996 is:



Shown alternatively:



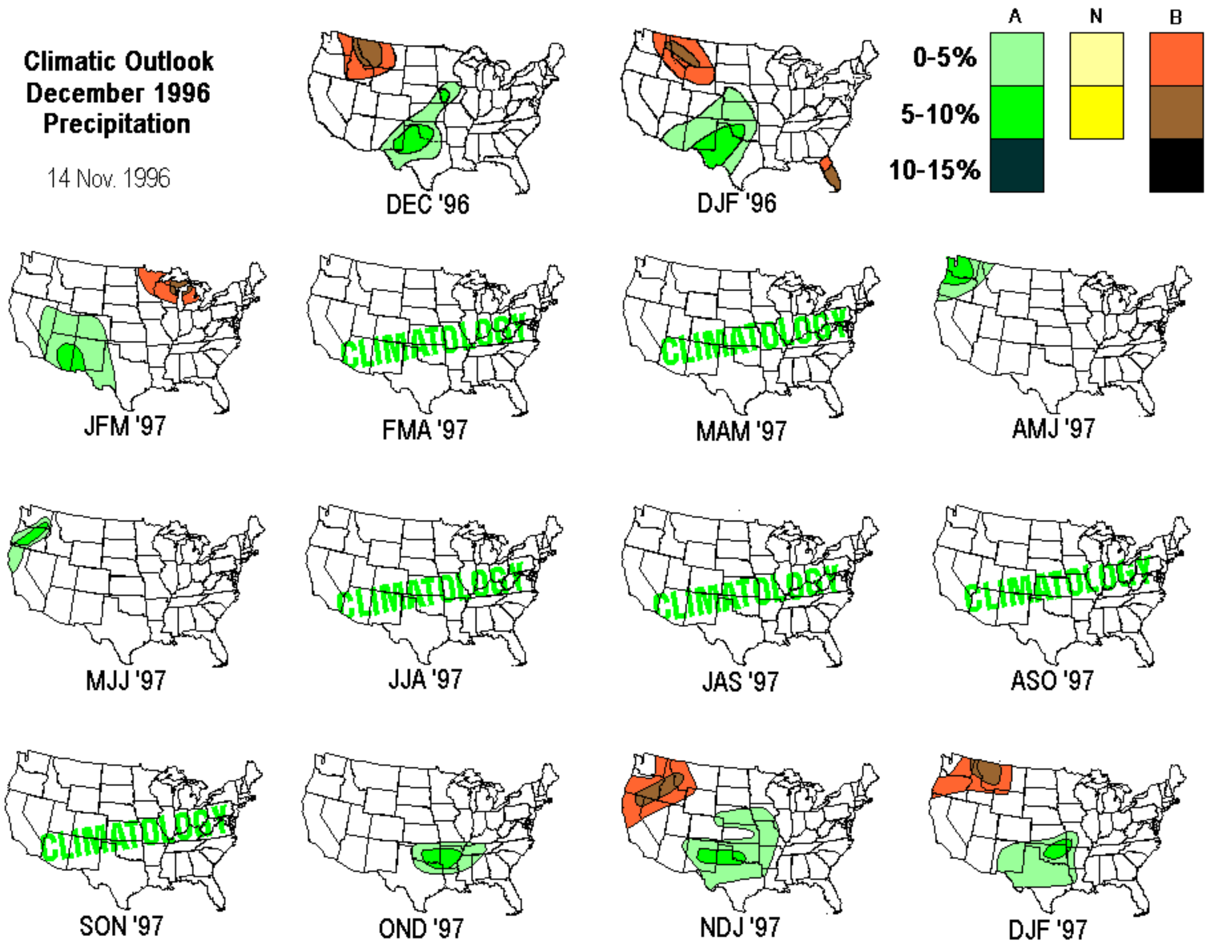
Forecast probabilities can be ascertained for any point on the outlook map. For example, at the asterisk, the probability of December precipitation in the upper third of historical observations is 0.383, the probability of December precipitation in the middle third of historical observations is 0.333, and the probability of December precipitation in the lower third of historical observations is 0.283.

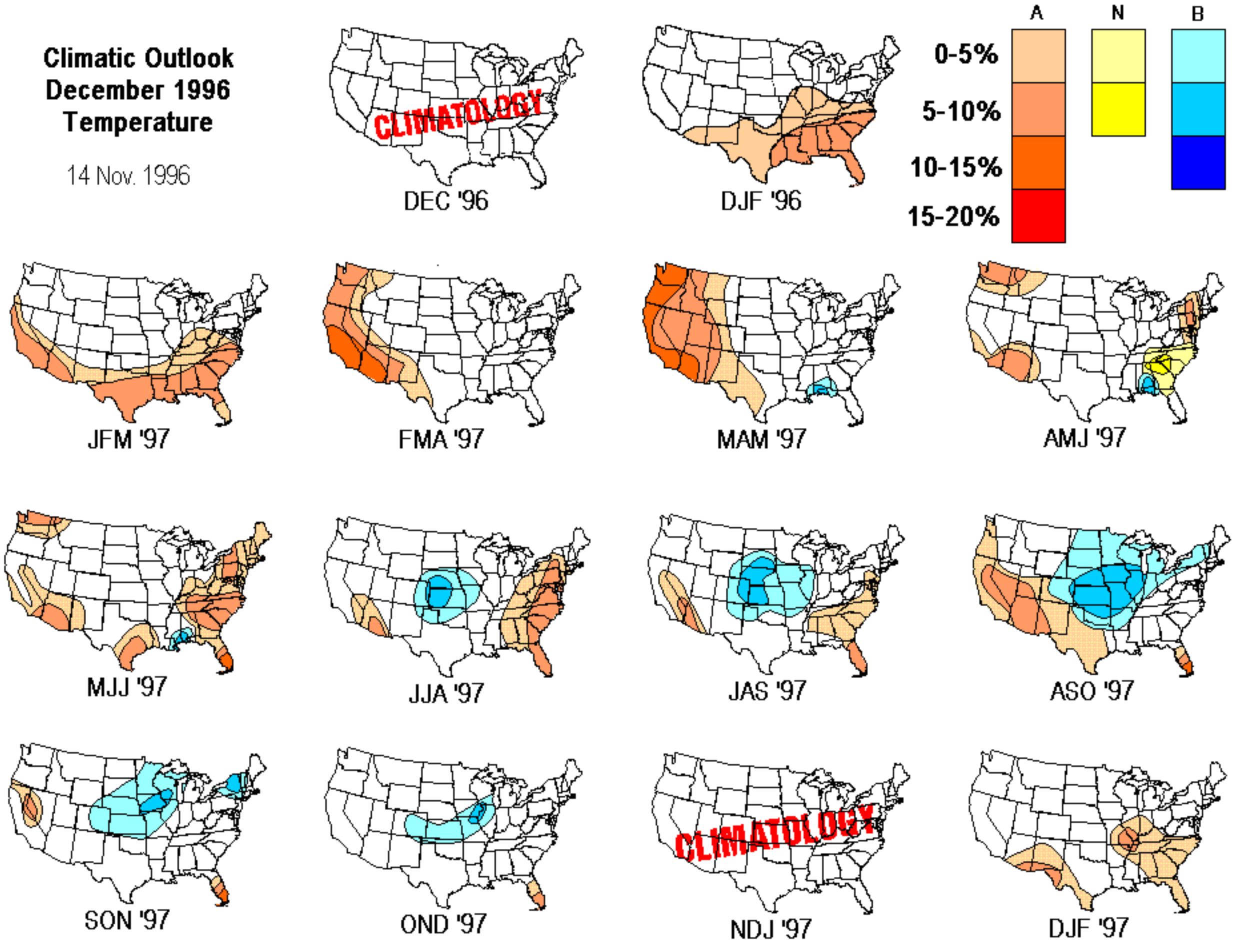
$$P[p_{Dec,0.667} < P_{Dec}] = 0.383$$

~~$P[p_{Dec,0.333} < P_{Dec} \le p_{Dec,0.667}] = 0.333$~~ (redundant)

$$P[P_{Dec} \le p_{Dec,0.333}] = 0.283$$

Actually, there are multiple 1-month & 3-month outlooks every month from NOAA; for example, for December 1996:





These represent a large number of probability statements. For example, the multiple outlooks for December 1996 consist of one 1-month and 13 three-month outlooks (14) of both temperature and precipitation (x 2), each with 3 equations (1 redundant) (x 2) for a total of 56 equations per monthly NOAA climatic outlook. Example numbers are taken from the maps for the Lake Superior basin.

Event Probabilities
December 1996 Air Temperature & Precipitation
DJF 1996 through DJF 1997 Air Temperature & Precipitation
forecast 14 November 1996 by NOAA:

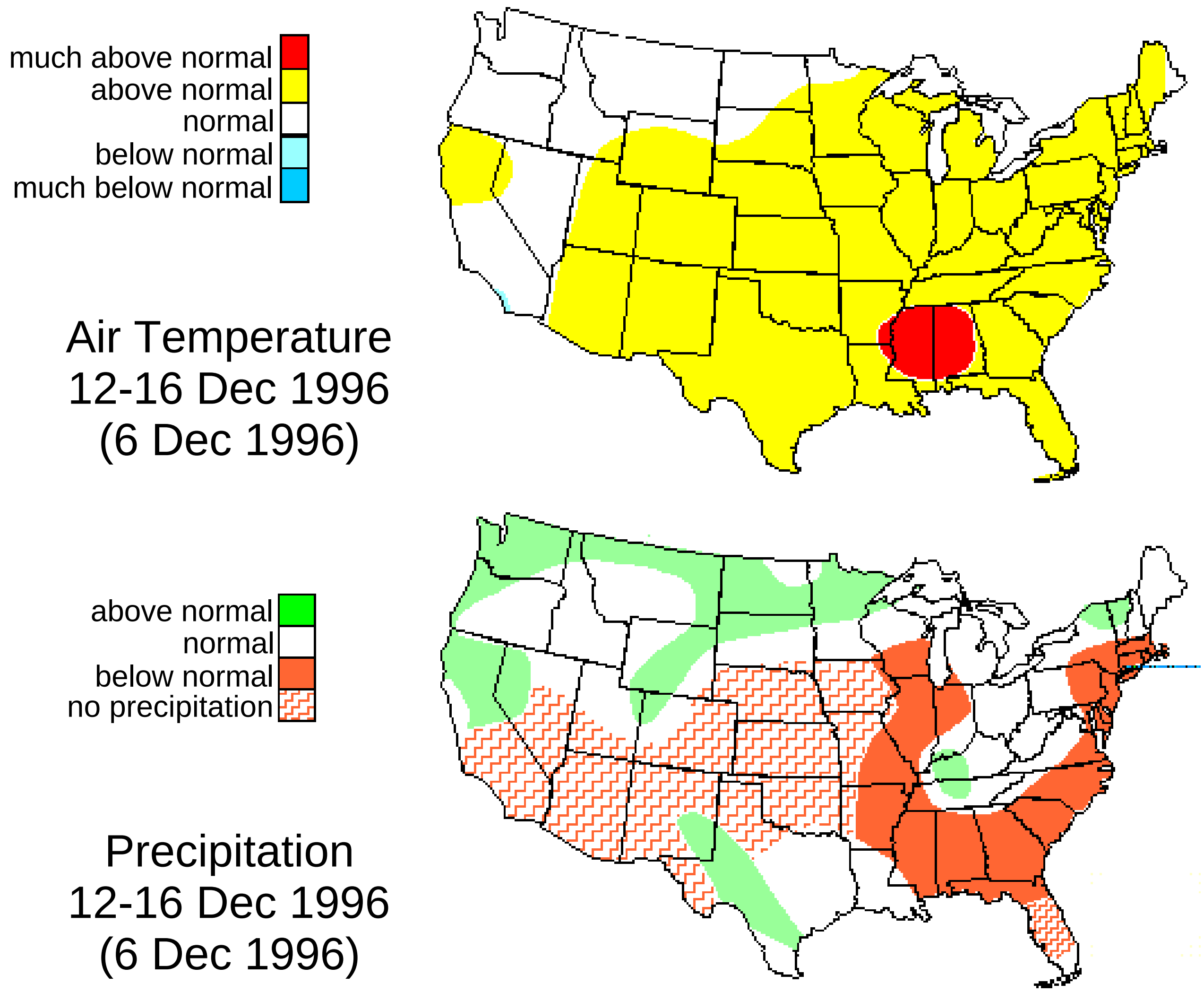
$$\hat{P}\left[T_{Dec96} \le t_{Dec,\,0.333}\right] = 0.333$$
$$\hat{P}\left[T_{Dec96} > t_{Dec,\,0.667}\right] = 0.333$$
$$\hat{P}\left[Q_{Dec96} \le q_{Dec,\,0.333}\right] = 0.333$$
$$\hat{P}\left[Q_{Dec96} > q_{Dec,\,0.667}\right] = 0.333$$
$$\hat{P}\left[T_{DJF96} \le t_{DJF,\,0.333}\right] = 0.333$$
$$\hat{P}\left[T_{DJF96} > t_{DJF,\,0.667}\right] = 0.333$$
$$\hat{P}\left[Q_{DJF96} \le q_{DJF,\,0.333}\right] = 0.333$$
$$\hat{P}\left[Q_{DJF96} > q_{DJF,\,0.667}\right] = 0.333$$
$$\hat{P}\left[T_{JFM97} \le t_{JFM,\,0.333}\right] = 0.333$$
$$\hat{P}\left[T_{JFM97} > t_{JFM,\,0.667}\right] = 0.333$$
$$\hat{P}\left[Q_{JFM97} \le q_{JFM,\,0.333}\right] = 0.283$$
$$\hat{P}\left[Q_{JFM97} > q_{JFM,\,0.667}\right] = 0.383$$
$$\hat{P}\left[T_{FMA97} \le t_{FMA,\,0.333}\right] = 0.333$$
$$\hat{P}\left[T_{FMA97} > t_{FMA,\,0.667}\right] = 0.333$$
$$\hat{P}\left[Q_{FMA97} \le q_{FMA,\,0.333}\right] = 0.333$$
$$\hat{P}\left[Q_{FMA97} > q_{FMA,\,0.667}\right] = 0.333$$
$$\hat{P}\left[T_{MAM97} \le t_{MAM,\,0.333}\right] = 0.333$$
$$\hat{P}\left[T_{MAM97} > t_{MAM,\,0.667}\right] = 0.333$$
$$\hat{P}\left[Q_{MAM97} \le q_{MAM,\,0.333}\right] = 0.333$$
$$\hat{P}\left[Q_{MAM97} > q_{MAM,\,0.667}\right] = 0.333$$
$$\hat{P}\left[T_{AMJ97} \le t_{AMJ,\,0.333}\right] = 0.333$$
$$\hat{P}\left[T_{AMJ97} > t_{AMJ,\,0.667}\right] = 0.333$$
$$\hat{P}\left[Q_{AMJ97} \le q_{AMJ,\,0.333}\right] = 0.333$$
$$\hat{P}\left[Q_{AMJ97} > q_{AMJ,\,0.667}\right] = 0.333$$
$$\hat{P}\left[T_{MJJ97} \le t_{MJJ,\,0.333}\right] = 0.333$$
$$\hat{P}\left[T_{MJJ97} > t_{MJJ,\,0.667}\right] = 0.333$$
$$\hat{P}\left[Q_{MJJ97} \le q_{MJJ,\,0.333}\right] = 0.333$$
$$\hat{P}\left[Q_{MJJ97} > q_{MJJ,\,0.667}\right] = 0.333$$
$$\hat{P}\left[T_{JJA97} \le t_{JJA,\,0.333}\right] = 0.333$$
$$\hat{P}\left[T_{JJA97} > t_{JJA,\,0.667}\right] = 0.333$$
$$\hat{P}\left[Q_{JJA97} \le q_{JJA,\,0.333}\right] = 0.333$$
$$\hat{P}\left[Q_{JJA97} > q_{JJA,\,0.667}\right] = 0.333$$
$$\hat{P}\left[T_{JAS97} \le t_{JAS,\,0.333}\right] = 0.333$$
$$\hat{P}\left[T_{JAS97} > t_{JAS,\,0.667}\right] = 0.333$$
$$\hat{P}\left[Q_{JAS97} \le q_{JAS,\,0.333}\right] = 0.333$$
$$\hat{P}\left[Q_{JAS97} > q_{JAS,\,0.667}\right] = 0.333$$
$$\hat{P}\left[T_{ASO97} \le t_{ASO,\,0.333}\right] = 0.363$$
$$\hat{P}\left[T_{ASO97} > t_{ASO,\,0.667}\right] = 0.303$$
$$\hat{P}\left[Q_{ASO97} \le q_{ASO,\,0.333}\right] = 0.333$$
$$\hat{P}\left[Q_{ASO97} > q_{ASO,\,0.667}\right] = 0.333$$
$$\hat{P}\left[T_{SON97} \le t_{SON,\,0.333}\right] = 0.343$$
$$\hat{P}\left[T_{SON97} > t_{SON,\,0.667}\right] = 0.323$$
$$\hat{P}\left[Q_{SON97} \le q_{SON,\,0.333}\right] = 0.333$$
$$\hat{P}\left[Q_{SON97} > q_{SON,\,0.667}\right] = 0.333$$
$$\hat{P}\left[T_{OND97} \le t_{OND,\,0.333}\right] = 0.333$$
$$\hat{P}\left[T_{OND97} > t_{OND,\,0.667}\right] = 0.333$$
$$\hat{P}\left[Q_{OND97} \le q_{OND,\,0.333}\right] = 0.333$$
$$\hat{P}\left[Q_{OND97} > q_{OND,\,0.667}\right] = 0.333$$
$$\hat{P}\left[T_{NDJ97} \le t_{NDJ,\,0.333}\right] = 0.333$$
$$\hat{P}\left[T_{NDJ97} > t_{NDJ,\,0.667}\right] = 0.333$$
$$\hat{P}\left[Q_{NDJ97} \le q_{NDJ,\,0.333}\right] = 0.333$$
$$\hat{P}\left[Q_{NDJ97} > q_{NDJ,\,0.667}\right] = 0.333$$
$$\hat{P}\left[T_{DJF97} \le t_{DJF,\,0.333}\right] = 0.333$$
$$\hat{P}\left[T_{DJF97} > t_{DJF,\,0.667}\right] = 0.333$$
$$\hat{P}\left[Q_{DJF97} \le q_{DJF,\,0.333}\right] = 0.333$$
$$\hat{P}\left[Q_{DJF97} > q_{DJF,\,0.667}\right] = 0.333$$

Multiple Probabilistic Meteorology Outlooks

NOAA 6-10 Day Meteorology Outlooks

Multiple probabilistic meteorology outlooks can also consist of multiple forms of probability outlooks: event probabilities (such as the preceding NOAA CPC 1- & 3-Month Climatic Outlooks of Temperature & Precipitation) and most-probable events (examples are the NOAA CPC 6-10 Day Outlooks of Temperature & Precipitation).

NOAA 6-10 Day Outlooks



Most-probable event outlooks result in inequalities instead of equations! The NOAA CPC 6-10 Day Outlooks of Temperature & Precipitation result in 8 inequalities per outlook.

Most-Probable Event
12-16 Dec 1996 Air Temperature and Precipitation
forecast 6 December 1996 by NOAA:

$$\hat{P}\left[T_{12-16Dec96} \le t_{12-16Dec,\,0.100}\right] \le 0.100$$

$$\hat{P}\left[t_{12-16Dec,\,0.100} < T_{12-16Dec96} \le t_{12-16Dec,\,0.300}\right] \le 0.200$$

$$\hat{P}\left[t_{12-16Dec,\,0.300} < T_{12-16Dec96} \le t_{12-16Dec,\,0.700}\right] \le 0.400$$

$$\hat{P}\left[t_{12-16Dec,\,0.700} < T_{12-16Dec96} \le t_{12-16Dec,\,0.900}\right] > 0.200$$

$$\hat{P}\left[T_{12-16Dec96} > t_{12-16Dec,\,0.900}\right] \le 0.100$$

$$\hat{P}\left[Q_{12-16Dec96} \le q_{12-16Dec,\,0.333}\right] \le 0.333$$

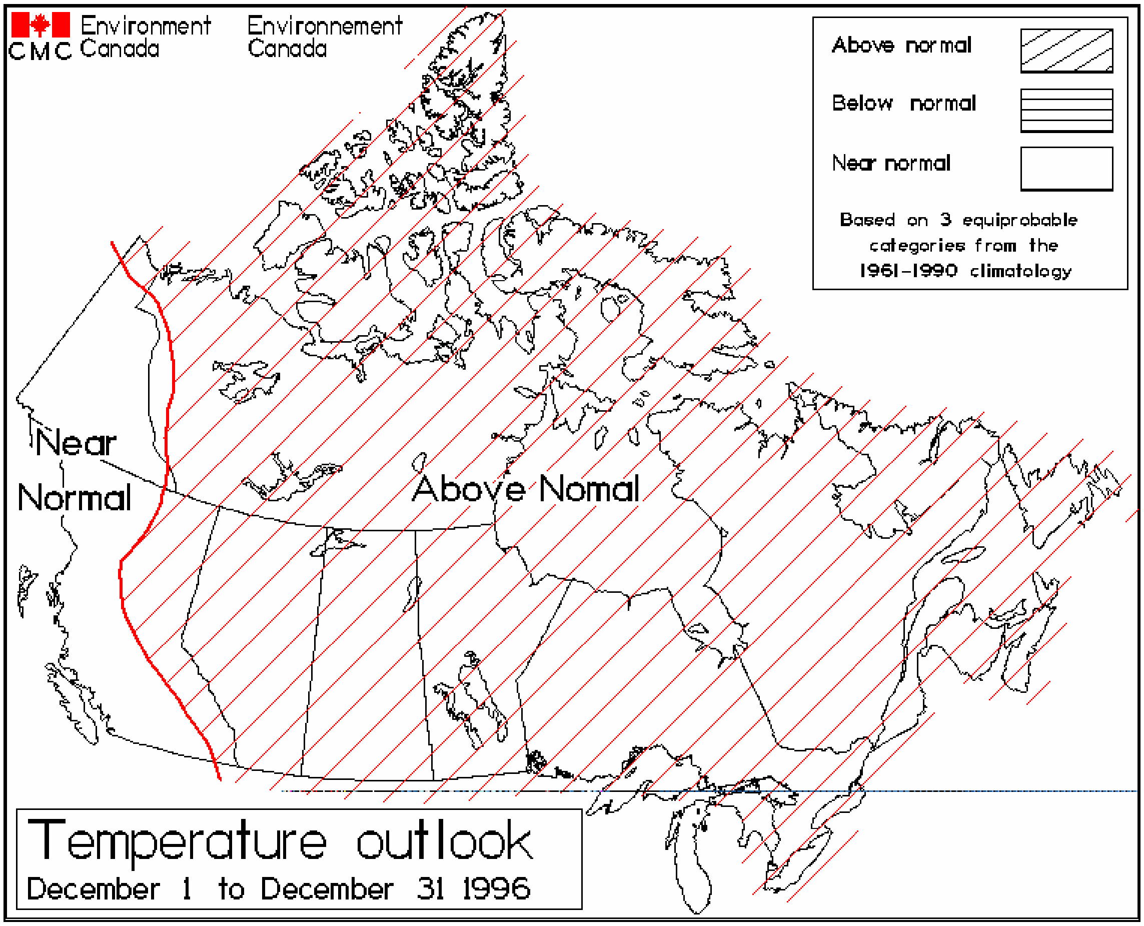
$$\hat{P}\left[q_{12-16Dec,\,0.333} < Q_{12-16Dec96} \le q_{12-16Dec,\,0.667}\right] \le 0.334$$

$$\hat{P}\left[Q_{12-16Dec96} > q_{12-16Dec,\,0.667}\right] > 0.333$$

Multiple Probabilistic Meteorology Outlooks

EC 1- & 3-Month Temperature Outlooks

Other agencies also issue most-probable event meteorology outlooks. The Environment Canada (EC) Climate Meteorology Center currently issue a 1- and 3-month climate outlook of air temperature. The 1-month outlook is issued twice a month, and the 3-month outlook is issued 4 times a year.

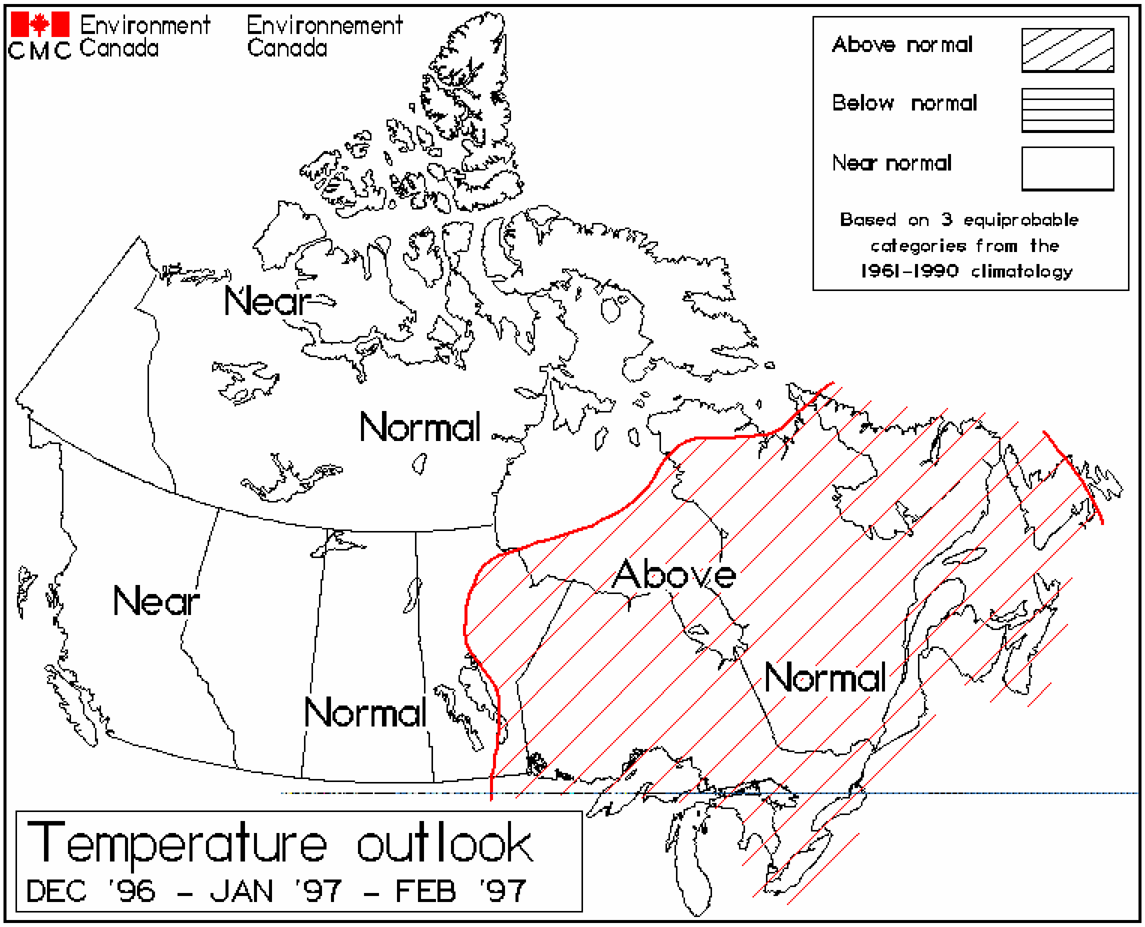


Most-Probable Event
Dec 1996 Air Temperature
forecast 1 Dec 1996 by EC:

$$\hat{P}\left[T_{Dec96} \le t_{Dec,\,0.333}\right] \le 0.333$$

$$\hat{P}\left[t_{Dec,\,0.333} < T_{Dec96} \le t_{Dec,\,0.667}\right] \le 0.334$$

$$\hat{P}\left[T_{Dec96} > t_{Dec,\,0.667}\right] > 0.333$$



Most-Probable Event
DJF 1996 Air Temperature
forecast 1 Dec 1996 by EC:

$$\hat{P}\left[T_{DJF96} \leq t_{DJF,\,0.333}\right] \leq 0.333$$

$$\hat{P}\left[t_{DJF,\,0.333} < T_{DJF96} \leq t_{DJF,\,0.667}\right] \leq 0.334$$

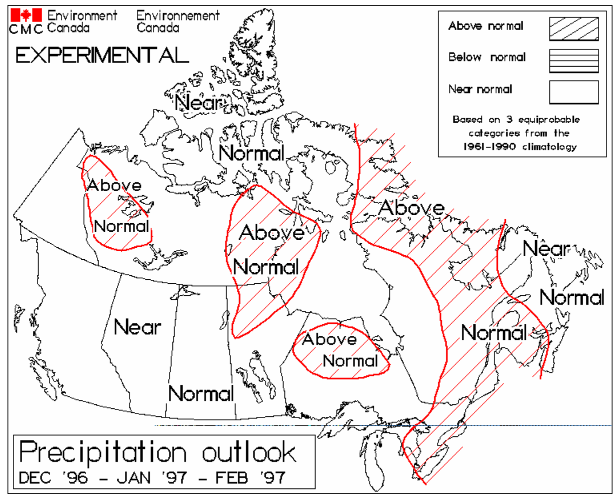
$$\hat{P}\left[T_{DJF96} > t_{DJF,\,0.667}\right] > 0.333$$

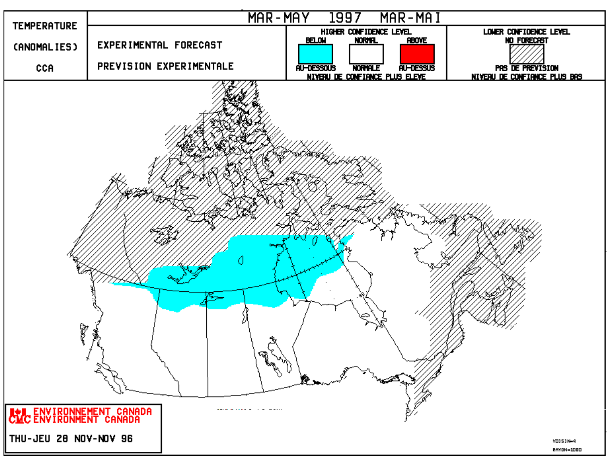
Multiple Probabilistic Meteorology Outlooks

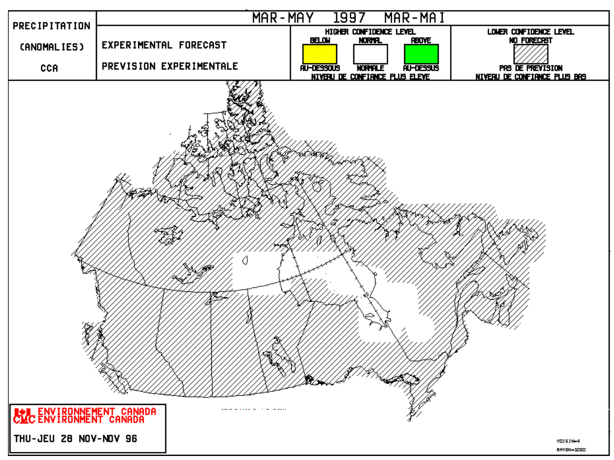
Pending Experimental Outlooks

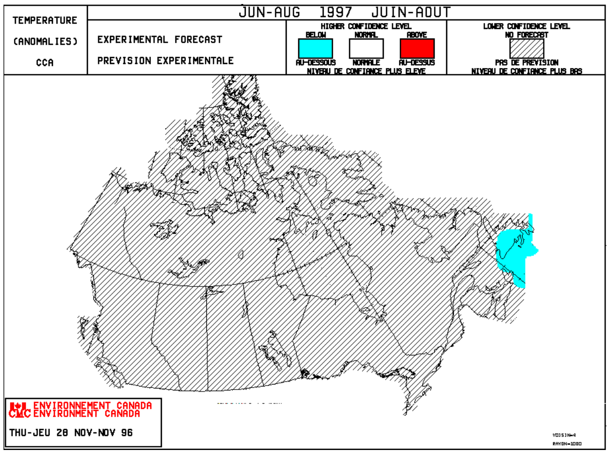
There are more to come! Environment Canada is currently considering experimentally another seven extended 3-month climate outlooks. These are comprised of a 3-month most-probable precipitation outlook over the next 3 months, 3 lagged (3, 6, & 9 months) 3-month most-probable temperature outlooks, and 3 lagged (3, 6, & 9 months) 3-month most-probable precipitation outlooks.

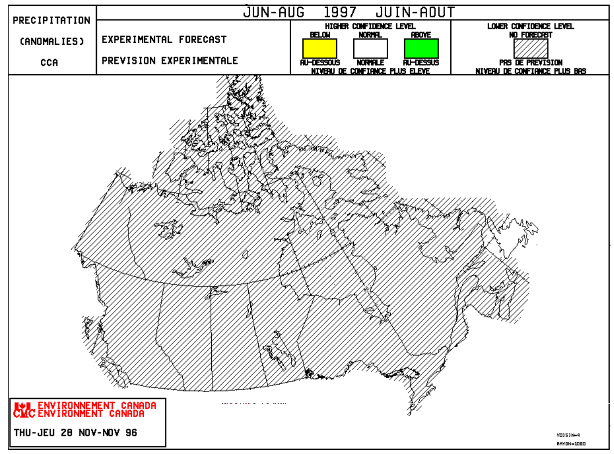
More experimental outlooks are currently under testing:

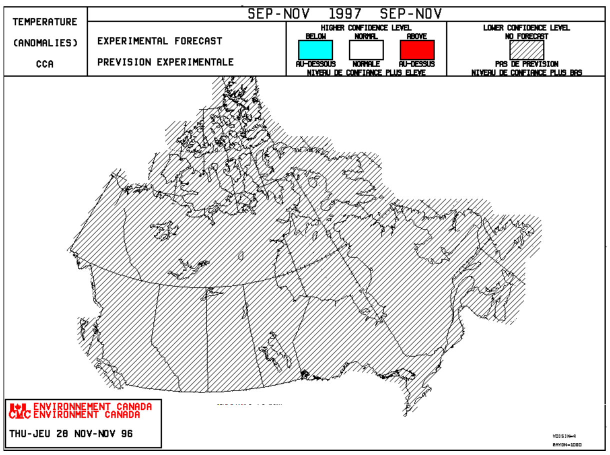


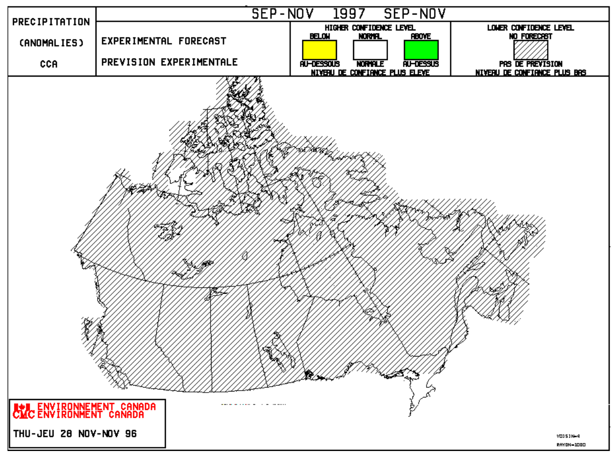












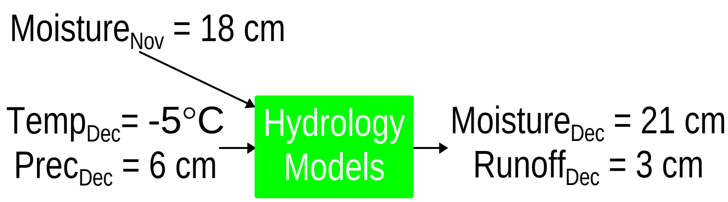
These 7 additional experimental outlooks will yield another 21 equations when they become operational!

Currently have 70 equations representing probabilistic meteorology outlooks. The outlooks are mixed event probabilities & most probable events. More are on the way.

Making A Hydrology Outlook

How do we make a hydrology outlook from a meteorology outlook? A first deterministic example, made 1 December for December, uses the following procedure:

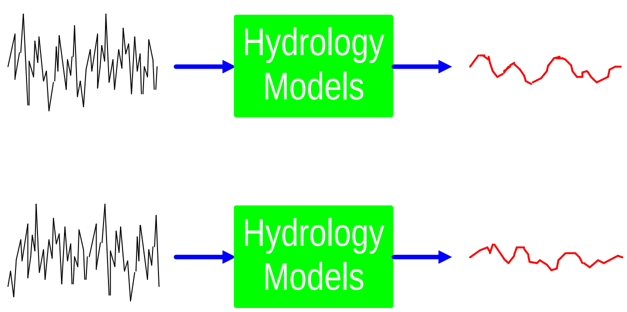
- Use the present hydrologic state as initial conditions in a model simulation.
- Transform the meteorology outlook to a hydrology outlook with models.



The disadvantage is that the hydrology outlook skill is tied to the low skill of the (deterministic) meteorology outlook.

A second deterministic example uses the following procedure:

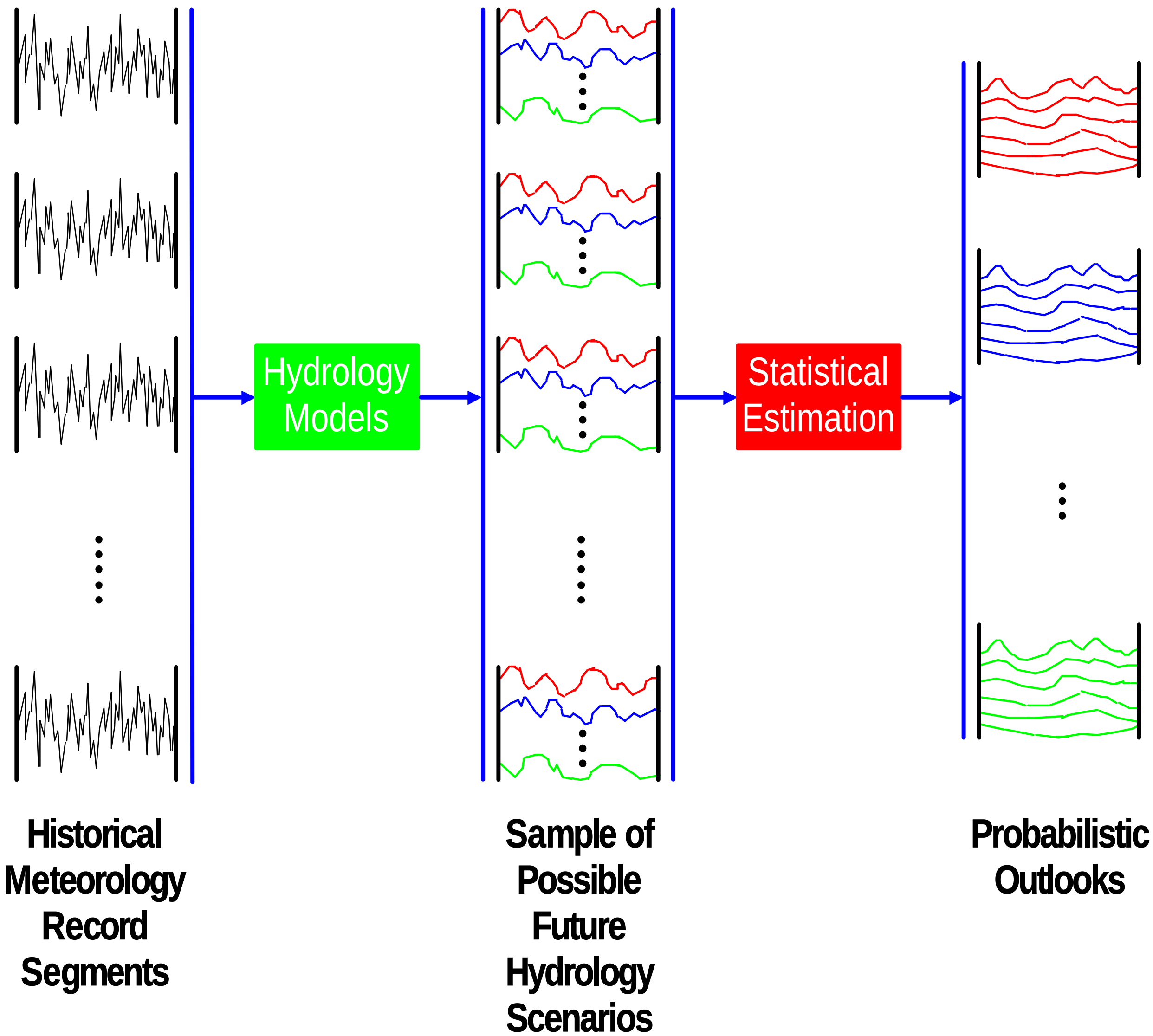
- Use the present hydrologic state as initial conditions in model simulations.
- Select segments of the historical record that resemble the meteorology outlook.
- Transform the record segments to possible hydrology scenarios with models.
- Select the "best" match from the multiple outlooks (either before or after the transformation).



The disadvantage is that there is still low skill in the meteorology outlook and the problem of selecting the "best" outlook.

An Operational Hydrology Approach

How do we make a hydrology outlook from a meteorology outlook (continued)? A first probabilistic example uses an *operational hydrology* approach:



- Use the present hydrologic state as initial conditions in model simulations.
- Select segments of the historical meteorology record.
- Transform record segments into hydrology scenarios with models. Each "scenario" has meteorology & hydrology variables for a single time period. The 1st "scenario" is produced from the 1st "historical meteorology record segment", the 2nd "scenario" produced from the 2nd "historical meteorology record segment", and so forth.
- Consider the set of possible scenarios (of meteorology & hydrology) as a statistical sample.

- Infer hydrology probabilities through statistical estimation by estimating probabilities with relative frequencies. For example, count the scenarios with December runoff less than or equal to 3 cm and divide by the total number of scenarios; count the scenarios with water surface temperature greater than 4°C and divide by the total:

$$\hat{P}[R_{Dec} \leq 3\,cm] = \frac{n(r_{Dec} \leq 3\,cm)}{n(\mathrm{all})}$$

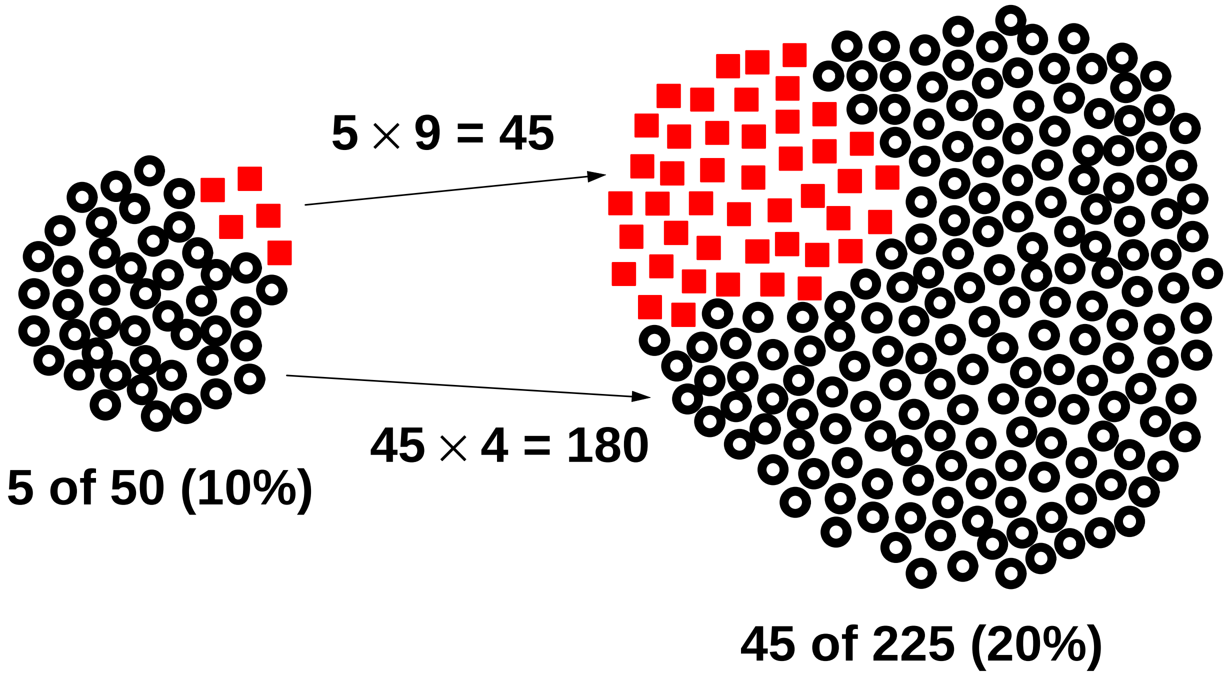
$$\hat{P}[W_{Dec} > 4^{\circ}\mathrm{C}] = \frac{n(w_{Dec} > 4^{\circ}\mathrm{C})}{n(\mathrm{all})}$$

The disadvantage is that it does not match the meteorology outlook.

Building A Structured Set

How do we make a hydrology outlook from a meteorology outlook (continued)? A second probabilistic example uses a biased operational hydrology approach (the first 3 steps are the same)

- Use the present hydrologic state as initial conditions in model simulations.
- Select segments of the historical meteorology record.
- Transform record segments to a set (sample) of hydrology scenarios with models.
- "Bias" the sample (use more of some scenarios than others) to match a meteorology outlook. Each "item" in the figure represents a scenario, consisting of both meteorology & modeled hydrology variables for a single time period.
  - ⇒ Duplicate each scenario in the sample by a duplication count.
  - ⇒ Choose all duplication counts so that the relative frequency of a desired event matches the meteorology outlook.
  - ⇒ For example, only 5 of 50 (10%) scenarios have a December temperature > 7°C but the meteorology outlook says that the probability is 20% that December temperature > 7°C. Repeat each of these 5 scenarios 9 times and repeat the other 45 scenarios 4 times to build a biased sample that has 45 of 225 (20%) with December temperature > 7°C.
  - ⇒ Of course, this duplication also applies to all variables in each scenario, both hydrology as well as meteorology.



- Infer hydrology probabilities through statistical estimation on the biased sample, by estimating probabilities with relative frequencies. For example

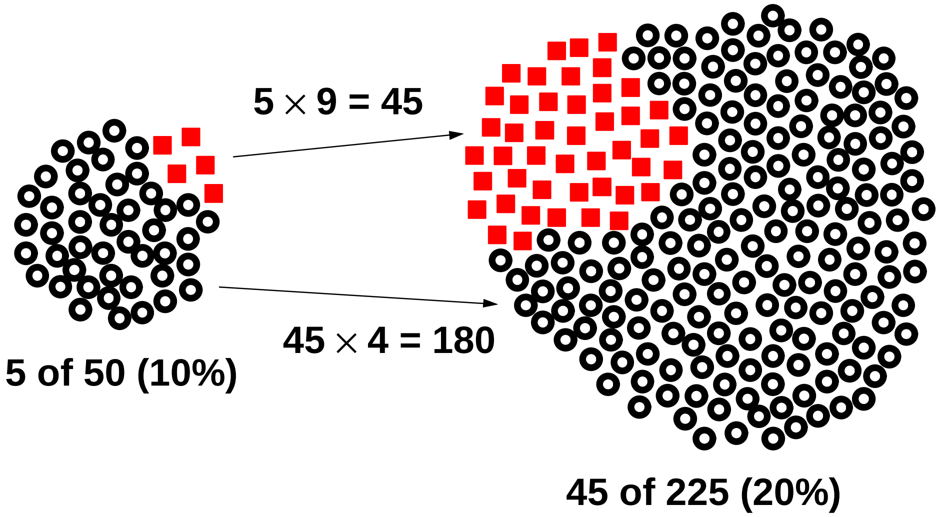
$$\hat{P}[R_{Dec} \le 3\,cm] = \frac{N(r_{Dec} \le 3\,cm)}{N(\text{all})}$$

$$\hat{P}[W_{Dec} > 4^{\circ}\text{C}] = \frac{N(w_{Dec} > 4^{\circ}\text{C})}{N(\text{all})}$$

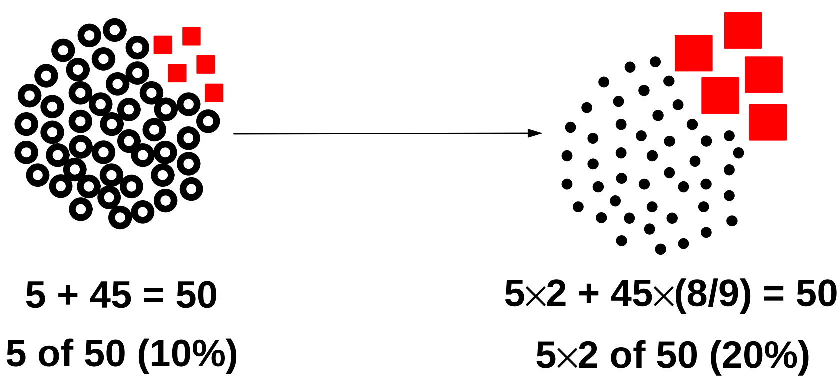
The advantages are that the biased methodology now matches the meteorology outlook and there is more skill in probabilistic meteorology outlooks than in deterministic. The disadvantage is that the method is only good for one meteorology outlook.

Reconsidering Resampling

How do we make a probabilistic hydrology outlook from multiple meteorology outlooks? First, note that for one setting we can express the resampling more generally in terms of "weights" applied to the original sample of scenarios. Its more convenient to work with the original sample of scenarios. For example, instead of duplicating five scenarios nine times and 45 scenarios four times to build a large sample of 225,



apply the weight of 2 to the five scenarios and 8/9 to the other 45 to get a weighted sample size of 50 with a weighted fraction of 20% of the desired scenarios.



This is simply another way of expressing the "biased" sample concept, but in terms of the original sample. It makes scenarios of interest a larger percentage of the sample. We then have a weight for each meteorology record piece and for the corresponding hydrology variables. We can again infer probabilities, by estimating probabilities with relative frequencies, expressed in terms of the sample weights.

$$\hat{P}[X \le x] = \frac{N\{x_k \le x\}}{N} = \frac{1}{n}\sum_{\{i | x_i \le x\}} w_i$$

We can similarly write an equation to match all 70 probabilistic meteorology outlooks. For the forecast event probabilities, there will be an equation of the above form, repeated here:

$$\text{For } \hat{P}[X \le x] = g, \qquad \frac{1}{n} \sum_{\{i \mid x_i \le x\}} w_i = g$$

For the forecast most-probable events, there are two other forms that the probability statements can take (actually more than two, but they can all be reduced to one of these forms).

$$\text{For } \hat{P}[z_1 < X \le z_2] \le g, \qquad \frac{1}{n} \sum_{\{i \mid z_1 < x_i \le z_2\}} w_i \le g$$

$$\text{For } \hat{P}[z_1 < X \le z_2] > g, \qquad \frac{1}{n} \sum_{\{i \mid \text{not } (z_1 < x_i \le z_2)\}} w_i < 1 - g$$

After forming a system of such equations, we can solve for the weights (see following).

Rewriting The Equations

The equations and inequalities matching all of the probabilistic meteorology outlooks may be more simply written:

$$\begin{aligned}
\sum_{i=1}^{n} a_{k,i} w_i &= e_k & k &= 1,\ldots, m-1 \\
\sum_{i=1}^{n} w_i &= n & (k &= m) \\
\sum_{i=1}^{n} a_{k,i} w_i &< e_k & k &= m+1,\ldots m+p \\
\sum_{i=1}^{n} a_{k,i} w_i &\le e_k & k &= m+p,\ldots m+p+q
\end{aligned}$$

where there are m-1 probabilistic equations, p "greater-than" probabilistic inequalities, and q "less-than-or-equal-to" probabilistic inequalities. The $a_{k,i}$ coefficients are 0 or 1 corresponding to the exclusion or inclusion, respectively, of each variable in the sets and e_k corresponds to the multiple settings of the climate outlooks. Also included in the set of equations is the requirement that all weights sum to the sample size n (guaranteeing all probabilities sum to unity). This set of equations and inequalities is equivalent to:

$$\begin{aligned}
\sum_{i=1}^{n+p+q} a_{k,i} w_i &= e_k & k &= 1,\ldots, m+p+q \\
w_i &> 0, & i &= n+1,\ldots n+p \\
w_i &\ge 0, & i &= n+p+1,\ldots n+p+q
\end{aligned}$$

where w_i (i = n+1, …, $n + p + q$) are "slack" variables, added to change consideration of an inequality to consideration of an equality, and the additional coefficients are:

$$\begin{aligned}
a_{k,i} &= 0, & k &= 1,\ldots, m, & i &= n+1,\ldots, n+p+q \\
&= 1, & k &= m+1,\ldots, m+p+q, & i &= n+k-m \\
&= 0, & k &= m+1,\ldots, m+p+q, & i &> n, \quad i \ne n+k-m
\end{aligned}$$

Solving The Equations

We have $m+p+q$ equations in $n+p+q$ variables to solve simultaneously. Generally, $m+p+q \neq n+p+q$ and some of the equations may be either redundant or non-intersecting with the rest and must be eliminated. Selection of some for elimination is facilitated by assigning each a priority reflecting its importance to the user. [The highest priority is given to the $m^{\underline{th}}$ equation guaranteeing that all relative frequencies sum to unity.]

Each equation, in priority order starting with the next-to-highest priority, is compared to the set of all higher-priority equations and eliminated if it is redundant or does not intersect the set. (Use Gauss-Jordan method of elimination to identify it as infeasible or redundant.) By starting with high priorities, each equation is compared with a known valid set of equations, and higher-priority equations are kept in preference to lower-priority ones. Thus we can always reduce the equation set so that $m+p+q \leq n+p+q$. If $m+p+q = n+p+q$, then the set can be solved directly via Gauss-Jordan elimination for the weights, w_i, since the equations are now independent and intersecting ($M = N = m+p+q = n+p+q$ below).

- $y_j = j, \quad j = 1, \ldots, N$ ←variable j is in column j initially
- $z_i = i, \quad i = 1, \ldots, M$ ←equation i is in row i initially
- $i = 1$ ←start with row 1
- WHILE $i \leq M$
  - $j = i$
  - WHILE $j \leq N$ & $a_{z_i, y_j} = 0$ ←find (next) variable j
    - $j = j + 1$
  - IF $j \leq N$ THEN ←if feasible and not redundant (j found)
    - IF $j > i$ THEN ←rotate columns i through j
      - $x = y_j$
      - $y_k = y_{k-1}, \quad k = j, j-1, \ldots, i+1$
      - $y_i = x$
    - LOOP $j = 1, \ldots, M$ ←eliminate variable i from other rows
      - IF $i \neq j$ & $a_{z_j, y_i} \neq 0$ THEN
        - $r = a_{z_j, y_i} / a_{z_i, y_i}$
        - $a_{z_j, k} = a_{z_j, k} - r\, a_{z_i, k}, \quad k = 1, \ldots, N+1$
  - ELSE ←eliminate infeasible or redundant row
    - $M = M - 1$
    - $z_j = z_{j+1}, \quad j = 1, \ldots, M$
    - $i = i - 1$
  - $i = i + 1$ ←prepare to redo new current row

Gauss—Jordon Elimination Method on Matrix *a*, with *M* Rows and *N+1* Columns

[If a row is recognized as redundant or infeasible, it is removed from the matrix. Since computations start with row 1, earlier rows are automatically given precedence.]

Weight Determination Optimization

There are multiple solutions for $m+p+q < n+p+q$, and the identification of the “best” set of weights requires the specification of a measure for comparing the solutions. One such measure is the deviation of weights from unity. We can formulate an optimization problem to minimize this measure in selecting a solution.

$$\min \sum_{i=1}^{n} \left(w_i - 1\right)^2 \quad \text{subject to}$$

$$\text{s.t.} \quad \sum_{i=1}^{n+p+q} a_{k,i} w_i = e_k \qquad k = 1, \ldots, m+p+q$$

We can convert this to a set of linear equations amenable to classical differential calculus by defining the “Lagrangian” function (where λ_k = the unit penalty of violating the $k^{\underline{th}}$ constraint in the optimization).

$$L = \sum_{i=1}^{n} \left(w_i - 1\right)^2 - \sum_{k=1}^{m+p+q} \lambda_k \left[\sum_{i=1}^{n+p+q} a_{k,j} w_i - e_k \right]$$

Setting the first derivatives to zero,

$$\frac{\partial L}{\partial w_i} = 2\left(w_i - 1\right) - \sum_{k=1}^{m+p+q} \lambda_k a_{k,i} = 0, \qquad i = 1, \ldots, n$$

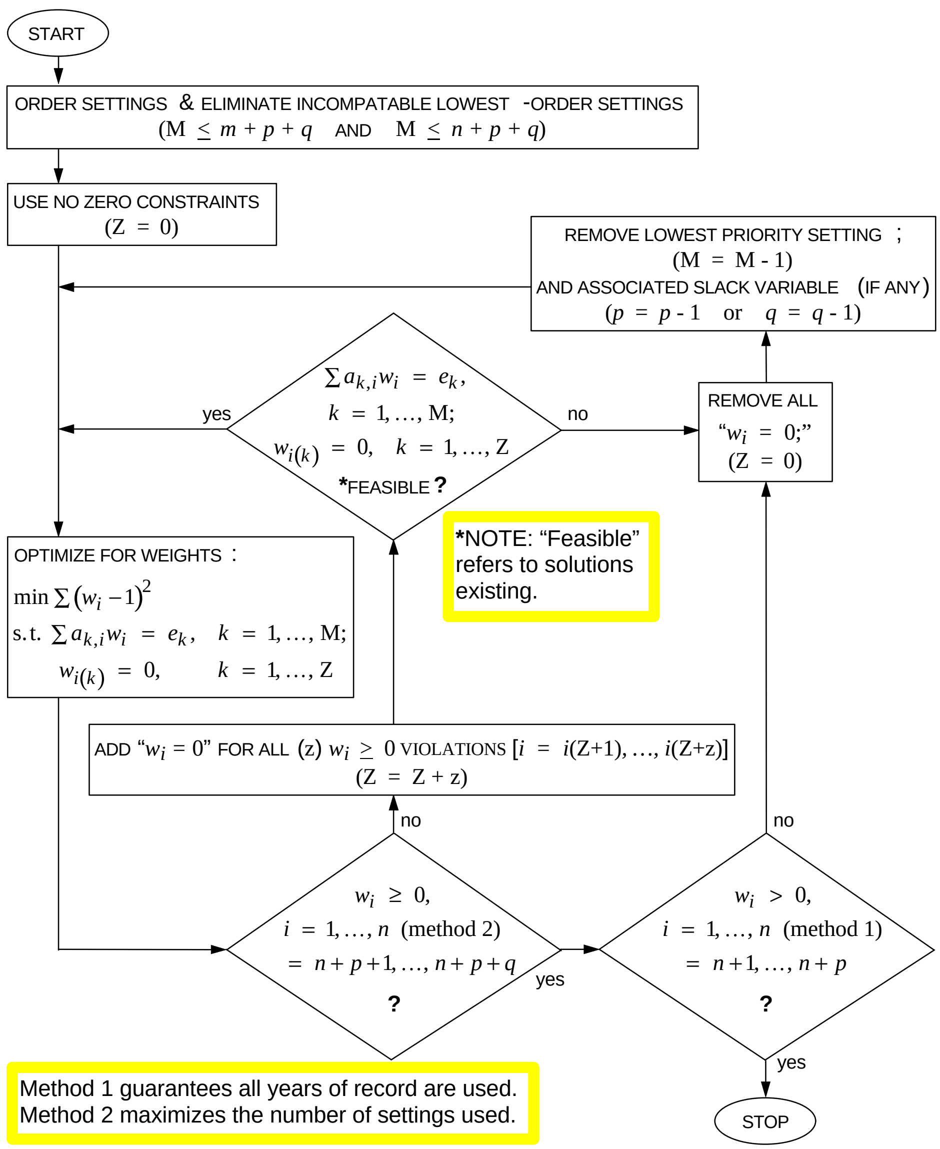
$$\frac{\partial L}{\partial w_i} = -\sum_{k=1}^{m+p+q} \lambda_k a_{k,i} = 0, \qquad i = n+1, \ldots, n+p+q$$

$$\frac{\partial L}{\partial \lambda_k} = -\sum_{i=1}^{n} a_{k,i} w_i + e_k = 0, \qquad k = 1, \ldots, m+p+q$$

These are necessary but not sufficient conditions for weight determination. They may be solved via the Gauss-Jordan method of elimination, because there are $m+n+2p+2q$ equations in $m+n+2p+2q$ unknowns, and sufficiency checked by inspection.

Optimization Algorithms

The solution may give positive, zero, or negative weights, but only non-negative weights make physical sense and we must further constrain the optimization by introducing non-negativity inequality constraints. The additional equations would require enumeration of all "zero points" or "roots" (a root is a solution with zero-valued weights). However, this is impractical since it can involve the inspection of many roots [e.g., for $n = 50$, there are $2^{50} - 1$ roots ($> 10^{15}$)]. Furthermore, non-negativity constraints can result in infeasibility (there is no solution). In this case, additional lowest-priority equations must be eliminated to allow a non-negative solution. The two methods portrayed in the figure provide systematic procedures for finding non-negative weights through elimination of lowest-priority equations. They avoid direct use of non-negativity constraints thus avoiding inspection of the large number of roots that can result. The first method guarantees strictly positive weights; all possible scenarios are used. The second method allows zero weights allowing satisfaction of more of the apriori settings.



Using The Weights

$$\hat{P}[X \le x] = \frac{1}{n} \sum_{\{i | x_i \le x\}} w_i$$

We have a weight for each piece of the historical meteorology record and we can use each weight with the hydrology scenario modeled from that meteorology record piece. Use the weights in the above equation but in the reverse direction (have weights, solve for probabilities) to make probability statements for variables other than meteorology (the hydrology variables).

Actually, we can derive other weighted statistics by transforming selected statistics defined over the very large sample (size N) to those defined over the biased sample (size n). For example:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} w_i \, x_i$$

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} w_i \left(x_i - \bar{x} \right)^2$$

$$\hat{P}\left[X \le y_j^n\right] = \frac{1}{n} \sum_{l=1}^{j} w_{i(l)}, \qquad j = 1, \ldots, n$$

where the estimate of the mean, m, is $\bar{x}$; the estimate of the variance, s^2, is S^2; and the estimate of the cumulative distribution function (non- exceedance probabilities), $P[X<x_g]=g$, is $\hat{P}\left[X \le y_j^n\right]$. Upper case X is any variable and lower case denotes values, x_i denotes the i^{th} sample value in the sample of size n, $\bar{x}$ and S^2 are the sample mean and variance of X, $\hat{P}[\;]$ is the “relative frequency” of an event in the sample, used as a probability estimate, and y_j^n is the j^{th} order statistic from smallest to largest. Note, $i(l)$ is the number of the value in the sample corresponding to the l^{th} order (for example, if the third value in the sample was the largest, then $i(n)$ = 3).

Example Consideration of Multiple Outlooks

The NOAA Climate Outlook for May 1997 (made 17 April 1997) over the Lake Ontario Basin is given here as three-digit probabilities.

Period, g	$\hat{P}\left[T_g \le t_{g,0.333}\right]$	$\hat{P}\left[T_g > t_{g,0.667}\right]$	$\hat{P}\left[Q_g \le q_{g,0.333}\right]$	$\hat{P}\left[Q_g > q_{g,0.667}\right]$
May '97	**0.333**	**0.333**	**0.333**	**0.333**
MJJ '97	**0.313**	**0.353**	**0.333**	**0.333**
JJA '97	**0.273**	**0.393**	**0.333**	**0.333**
JAS '97	**0.333**	**0.333**	**0.333**	**0.333**
ASO '97	**0.333**	**0.333**	**0.333**	**0.333**
SON '97	*0.383*	*0.283*	*0.333*	*0.333*
OND '97	*0.333*	*0.333*	*0.333*	*0.333*
NDJ '97	*0.333*	*0.333*	*0.333*	*0.333*
DJF '97	*0.303*	*0.363*	*0.333*	*0.333*
JFM '98	*0.233*	*0.433*	*0.333*	*0.333*
FMA '98	*0.293*	*0.373*	*0.333*	*0.333*
MAM '98	*0.333*	*0.333*	*0.333*	*0.333*
AMJ '98	*0.323*	*0.343*	*0.333*	*0.333*
MJJ '98	*0.263*	*0.403*	*0.333*	*0.333*

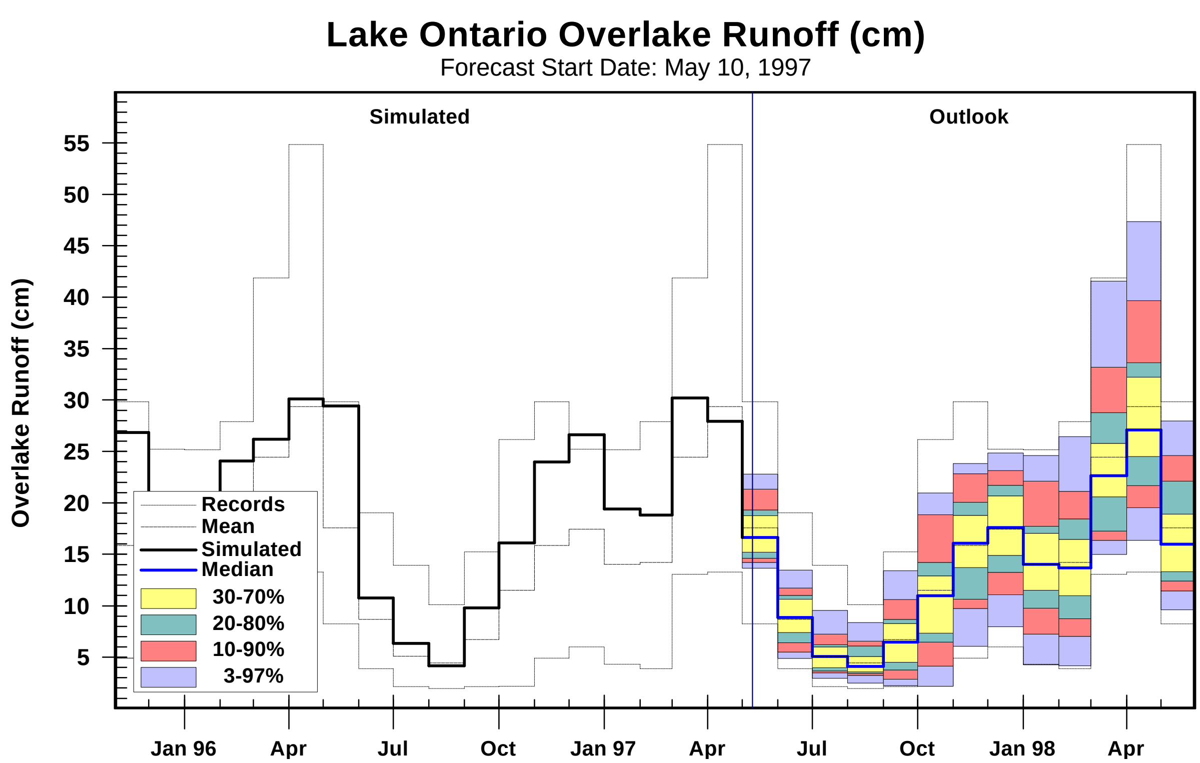
The first 20 values are used arbitrarily, in priority of their appearance, to make an outlook beginning 10 May 1997. They are used with the forty-seven 13-month time series, May-May, from the historical record of 1948-1995, to construct these 21 equations.

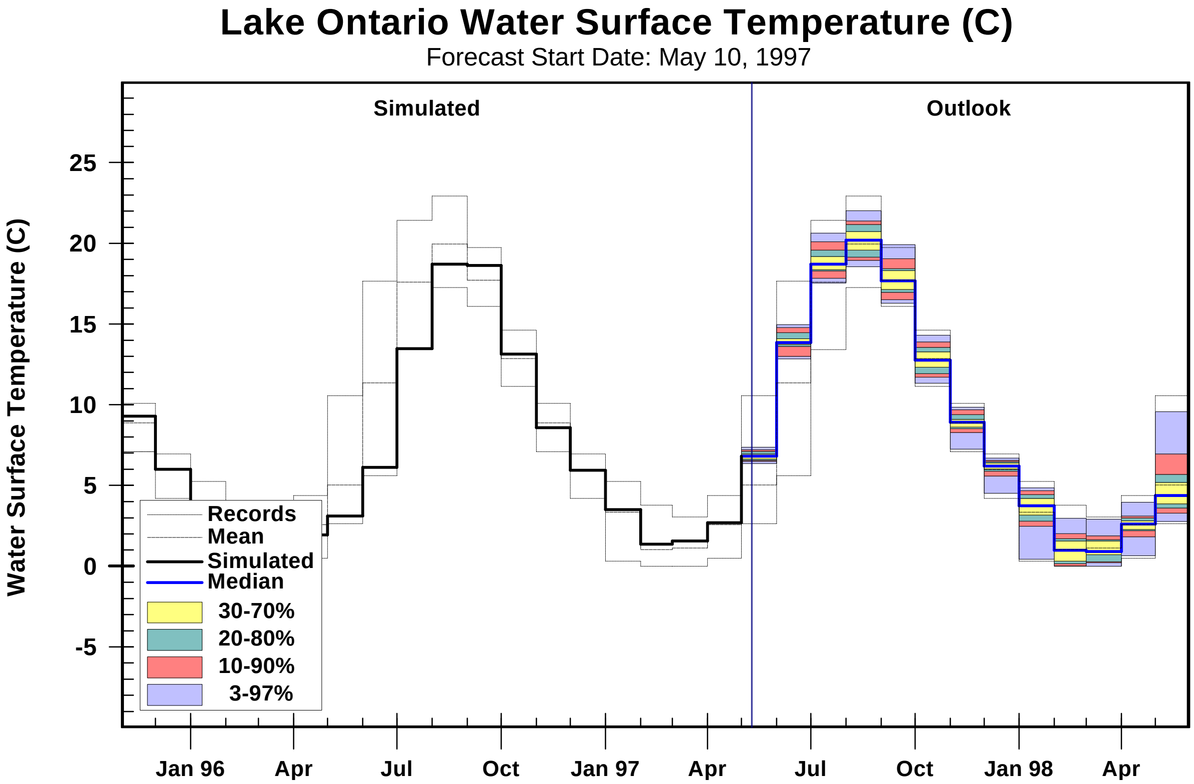
Period	k	Inclusion in interval, $a_{k,i}$, $i = 1, ..., 47$	e_k
Entire	1	11111111111111111111111111111111111111111111111	1.000×47
May '97	2	10001000101001010011100000101000000110000000001	0.333×47
May '97	3	00000001000110101100000000010100101000111001000	0.333×47
May '97	4	01110011001000100110000100000110100000011000010	0.333×47
May'97	5	10001100110010000000110011101000000110000110101	0.333×47
MJJ '97	6	00000000101001000101110000101000000011000000100	0.313×47
MJJ '97	7	01011101010100101000001001010100000000111101011	0.353×47
MJJ '97	8	011000110000001111100000000000111000000001001000	0.333×47
MJJ '97	9	00000000010001000001010010101000000010100110100	0.333×47
JJA '97	10	00110000101010100100100110000100001001100000100	0.273×47
JJA '97	11	1100110100010000000110000010100000000110011111011	0.393×47
JJA '97	12	010011010000101101100000001000011000100001001000	0.333×47
JJA '97	13	00010000000001000001000010011100110010100000100	0.333×47
JAS '97	14	00110010110000111101000000101000001000100000100	0.333×47
JAS '97	15	11001101000101000000111101000000100100011001010	0.333×47
JAS '97	16	10000000010011011010010001000000000100000110000	0.333×47
JAS '97	17	000100001010000000001001000011101010010110000100	0.333×47
ASO '97	18	00000000000000001010000010101100011000010000100	0.333×47
ASO '97	19	11010111000111000000111101000000000111000011000	0.333×47
ASO '97	20	11111100110011011010010101100000001000000001001	0.333×47
ASO '97	21	00000011001100000101000000011101010001100010100	0.333×47

The solution of the equations is found by minimizing the deviation of weights from unity (least squares) by utilizing all 20 climate outlook settings (method 2).

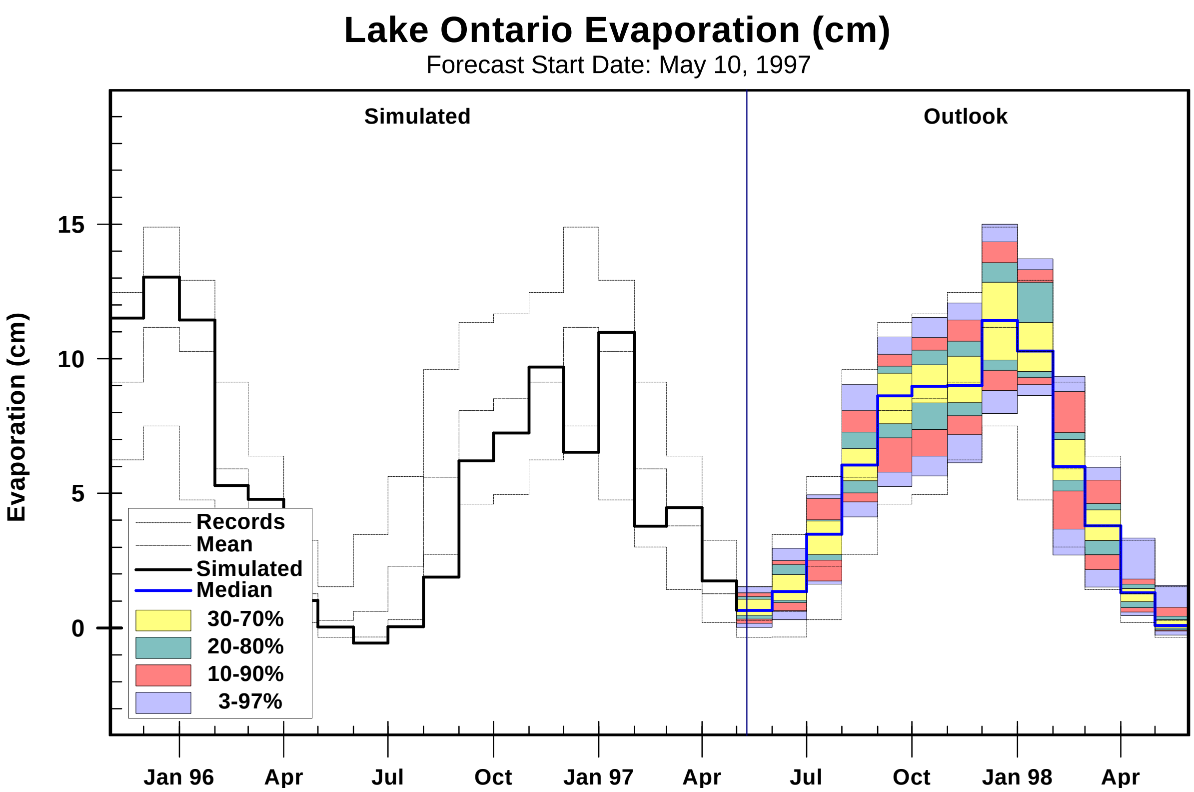
Year	Weight	Year	Weight	Year	Weight
1948	0.406380	1964	1.938431	1980	1.064842
1949	0.546473	1965	1.427033	1981	1.500636
1950	0.079226	1966	2.371934	1982	1.223223
1951	0.435065	1967	1.791979	1983	1.805120
1952	0	1968	0.708062	1984	1.576153
1953	0	1969	1.101582	1985	1.044314
1954	0.390604	1970	1.009006	1986	0.817660
1955	0.736226	1971	0	1987	1.919722
1956	0.060203	1972	1.404484	1988	1.346417
1957	0.395780	1973	0.856093	1989	1.359614
1958	0.735838	1974	1.030894	1990	0.937913
1959	0.464055	1975	0.732202	1991	0.727565
1960	0.840087	1976	1.526370	1992	1.642269
1961	2.066304	1977	1.093037	1993	0.893154
1962	1.320500	1978	1.290928	1994	0
1963	1.571764	1979	0.810861		

Note that four weights were assigned a value of zero to enable this inclusion, implying that the scenarios starting in May 1952, 1953, 1971, and 1994 are unused. An example probabilistic outlook is shown for 10 May '97 — May '98. There were 47 values of each modeled monthly variable (runoff, water temp., & evaporation), corresponding to the 47 simulations. Each was weighted to compute probabilistic outlook statistics.





Lake Ontario Water Surface Temperature (C)
Forecast Start Date: May 10, 1997
Simulated
Outlook
Water Surface Temperature (C)
25
20
15
10
5
0
-5
Records
Mean
Simulated
Median
30-70%
20-80%
10-90%
3-97%
Jan 96
Apr
Jul
Oct
Jan 97
Apr
Jul
Oct
Jan 98
Apr



Lake Ontario Evaporation (cm)
Forecast Start Date: May 10, 1997
Simulated
Outlook
Evaporation (cm)
15
10
5
0
Records
Mean
Simulated
Median
30-70%
20-80%
10-90%
3-97%
Jan 96
Apr
Jul
Oct
Jan 97
Apr
Jul
Oct
Jan 98
Apr

A Few Considerations

This is an operational hydrology approach to match climate outlooks and use all or most historical information. The National Weather Service's Extended Streamflow Prediction (ESP) approach is similar. Some other approaches limit historical data to match outlooks or ignore outlooks altogether.

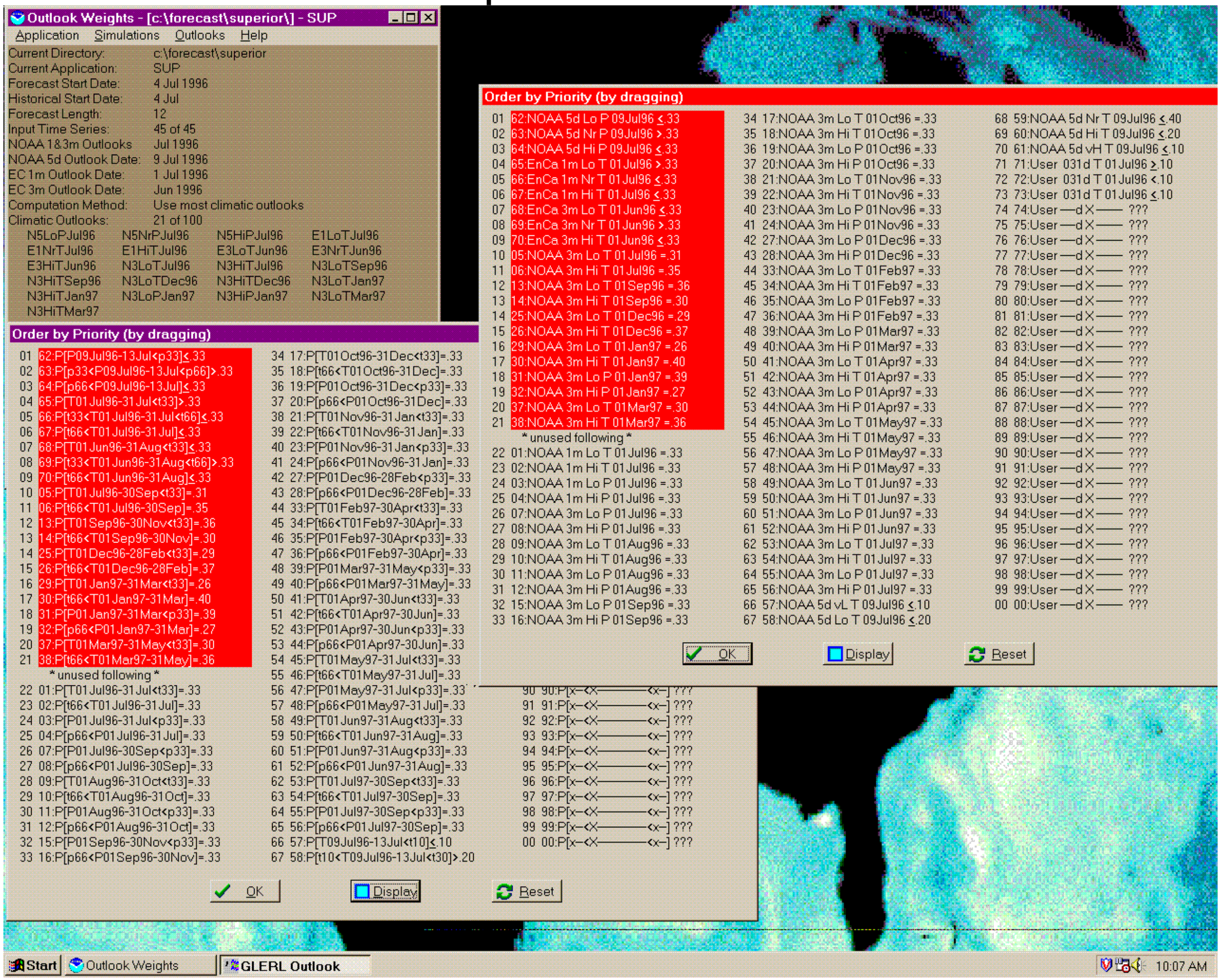
The use of a very large structured set of hydrology scenarios corresponds to a weighted original set, estimated from the historical record. ESP uses a similar approach estimating weights from climatic indices or (more recently) precipitation forecasts. Others use time series modeling; direct use of historical records avoids estimation problems but restricts the range of variables to historical.

Weight determination involves arbitrary choices at this point, such as the selection of an objective function. **min** $\sum(w_i - 1)^2$ strives toward all scenarios equally weighted. **min** $\sum w = n$ also strives toward equal weights but enables a linear programming solution; it is intractable because of the large number of roots but could be handled similarly. **min** $\sum[f_{\mathrm{hst}}(p, t) - f_{\mathrm{frcst}}(p, t)]^2$ forces joint distributions (before and after weighting) to be as similar as possible, but is intractable for multiple outlooks. Other objective functions, not yet formulated, could reflect additional consideration of climatic conditions or use climatic indices. Another arbitrary choice is in the selection of the climatic outlooks to use, the assignment of priorities to them, and even the priority structure suggested here.

An important advantage is the independence of the weights and the models. This is true with ESP also. After the hydrologic model simulations, many probabilistic outlooks can be generated, alternate weight formulations can be used, alternate climatic outlooks or outlook selections can be considered, alternate parametric approaches can be tried, and alternate methods of selecting climate outlooks to use (methods 1 and 2) can be investigated.

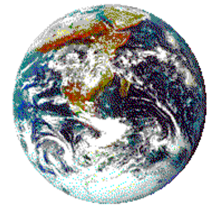
Code is available to utilize NOAA CPC 1- and 3-month air temperature and precipitation climate outlooks, NOAA CPC 6-10 day air temperature and precipitation climate outlooks, Environment Canada 1- and 3-month air temperature and pending precipitation climate outlooks, and user-defined probability outlooks.

Graphical User Interface



We built a graphical user interface as specially-designed *WindowsÔ*, *Windows95Ô* , and *Windows NTÔ* applications to allow ***<u>anyone</u>*** to directly use climate outlooks in their own applications. The climate outlooks considered are the NOAA 1- & 3-month event probabilities, the NOAA 6-10 day most-probable events, the Environment Canada 1- & 3-month most probable events, and user-defined outlook probability statements. This allows the user to set hydrologic outlook parameters and to begin a hydrological outlook. The interface defines the hydrological outlook and historical-data periods, selects the periods, probabilities, and priorities of climate outlooks, and determines the optimization method for considering the climate outlooks in making the hydrological outlook. The climate outlooks can be particularly cumbersome and difficult to use; but this interface greatly clarifies and simplifies their use in making a hydrological outlook. It allows readily understandable user interpretation of climate outlooks and easy user assignment of relevant priorities. This interface makes all computations utilizing the new climate outlooks. It finds all necessary reference quantiles for using a climate outlook from historical data, sets up all climate outlook selections as boundary equations and solves them (either by using all historical data or by maximizing use of the climate outlook selections). The interface computer code is also available as a stand-alone FORTRAN implementation for use under a variety of operating systems.

Probabilistic Hydrology Outlook Summary



Graphical User Interface for Windows™ & Windows95™

1. Read & enter all outlook meteorology probability statements.
2. Convert probability statements to "weights" equations.
   - a system of equations
   - solution yield weights to apply to hydrological scenarios
3. Eliminate redundant statements/equations.
4. Eliminate totally incompatible statements/equations.
5. Prioritize statements/equations.
   - may not be able to satisfy all statements/equations
   - so select order of importance

Automated Software Package

6. Determine operational hydrology parameters.
   - set forecast date
   - select pieces of historical record
   - set forecast length
7. Determine initial conditions for hydrology simulations.
   - acquire and organize most recent hydrometeorological data
   - use hydrology models with recent data to estimate present conditions
   - use available field measurements to set additional conditions
8. Simulate hydrology for all pieces of historical record.
   - use GLERL's hydrology models for rainfall/runoff, lake thermodynamics
   - or use hydrology models within the NWS River Forecast System
9. Solve equations to determine weights.
   - Croley, T. E., II (1996). "Using NOAA's new climate outlooks in operational hydrology." Journal of Hydrologic Engineering 1(3):93-102.
   - Croley, T. E., II (1997). "Water resource predictions from meteorological probability forecasts." In: Sustainability of Water Resources under Increasing Uncertainty (Proceedings Rabat Symposium, April 1997) (ed. By D. Rosbjerg et al.), IAHS Publication No. 240, IAHS Press, Wallingford, UK, 301-310.
   - Croley, T. E., II (1997). "Mixing probabilistic outlooks in operational hydrology." Journal of Hydrologic Engineering (in press, to appear Oct. 1997).
10. Apply weights to historical/modeled hydrometeorological "scenarios".

Back-End Graphical User Interface

11. Present probability distributions in graphical, tabular, file, and spreadsheet form(s).
12. Present for a variety of "dimensions."
   - 28 hydrometeorological variables
   - 7 lake surfaces
   - 121 watersheds
   - daily, weekly, quarter-monthly, monthly, annual