AREP
GAW

Section 12
Air Quality Forecasting Tools

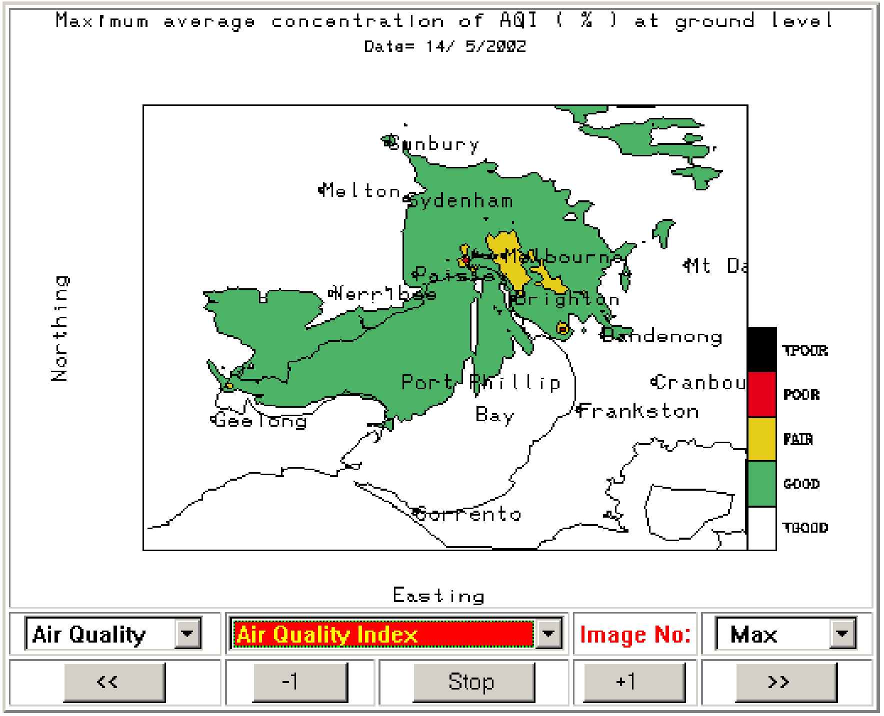
WMO
OMM

AREP
GAW

Background

- Forecasting tools provide information to help guide the forecasting process.
- Forecasters use a variety of data products, information, tools, and experience to predict air quality.
- Forecasting tools are built upon an understanding of the processes that control air quality.
- Forecasting tools:
  - Subjective
  - Objective
- More forecasting tools = better results.

Today's forecast-Melbourne



www.epa.vic.gov.au/air/AAQFS

WMO
OMM

Forecasting Tools and Methods (1 of 3)

- Persistence
- Climatology
- Criteria
- Statistical
  - Classification and Regression Tree (CART)
  - Regression
- Neural networks
- Numerical modeling
- Phenomenological and experience
- Predictor variables

Fewer resources, lower accuracy

More resources, potential for higher accuracy

Forecasting Tools and Methods (2 of 3)

Tool development is a function of

- Amount and quality of data (air quality and meteorological)
- Resources for development
  - Human
  - Software
  - Computing
- Resources for operations
  - Human
  - Software
  - Computing

Forecasting Tools and Methods (3 of 3)

For each tool

- What is it?
- How does it work?
- Example
- How to develop it?
- Strengths
- Limitations

Persistence (1 of 2)

- Persistence means to continue steadily in some state.
  - Tomorrow's pollutant concentration will be the same as Today's.
- Best used as a starting point and to help guide other forecasting methods.
- It should not be used as the only forecasting method.
- Modifying a persistence forecast with forecasting experience can help improve forecast accuracy.

Persistence forecast

Monday	Tuesday	Wednesday
Unhealthy	Unhealthy	Unhealthy

Persistence (2 of 2)

- Seven high ozone days (red)
- Five of these days occurred after a high day (*)
- Probability of high ozone occurring on the day after a high ozone day is 5 out of 7 days
- Probability of a low ozone day occurring after a low ozone day are 20 out of 22 days
- Persistence method would be accurate 25 out of 29 days, or 86% of the time

Peak 8-hr ozone concentrations for a sample city

Day	Ozone (ppb)	Day	Ozone (ppb)
1	80	16	**120***
2	50	17	**110***
3	50	18	80
4	70	19	80
5	80	20	70
6	**100**	21	60
7	**110***	22	50
8	**90***	23	50
9	80	24	70
10	80	25	80
11	80	26	80
12	70	27	70
13	80	28	80
14	**90**	29	60
15	**110***	30	70

Persistence – Strengths

- Persistence forecasting
  - Useful for several continuous days with similar weather conditions
  - Provides a starting point for an air quality forecast that can be refined by using other forecasting methods
  - Easy to use and requires little expertise

Persistence – Limitations

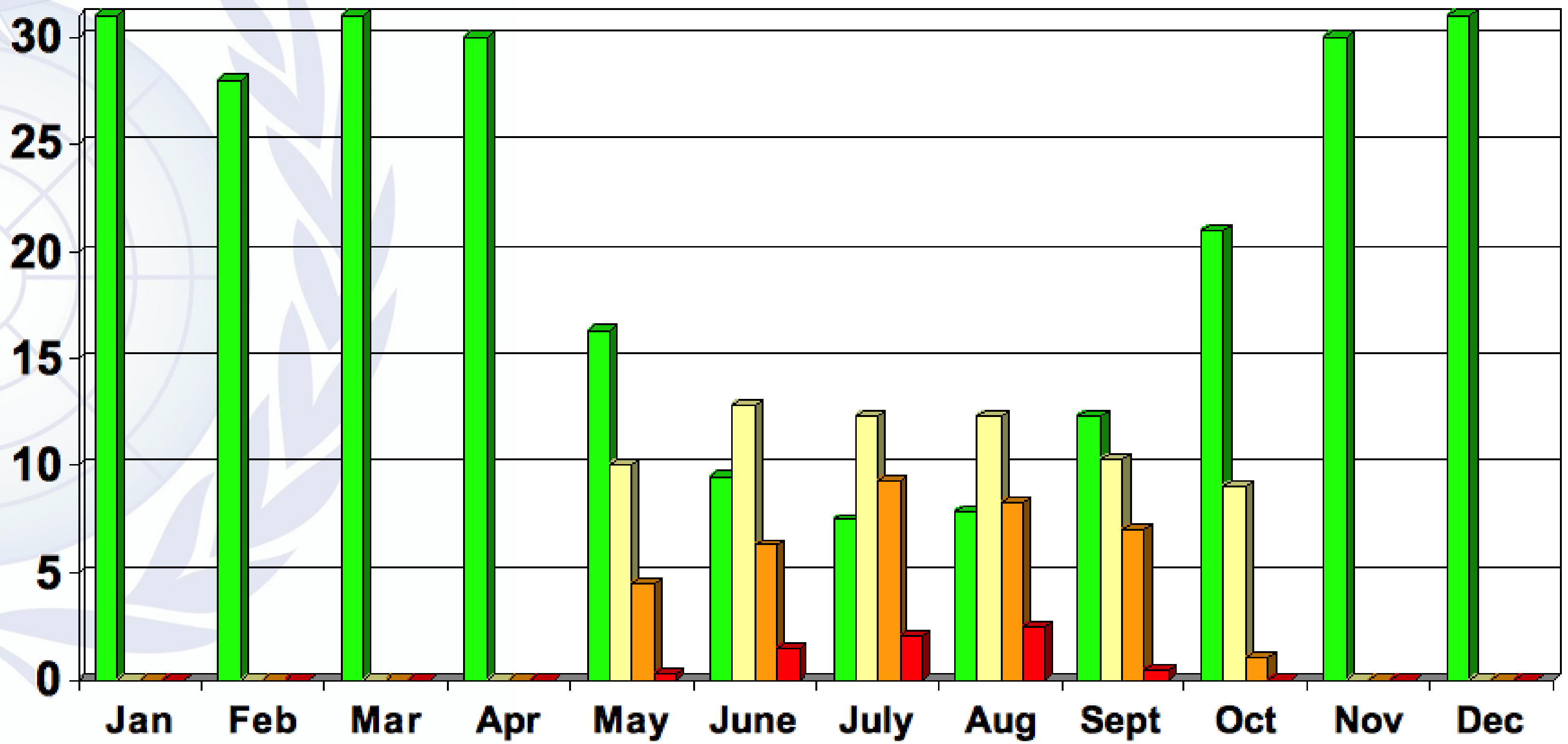
- Persistence forecasting cannot
  - Predict the start and end of a pollution episode
  - Work well under changing weather conditions when accurate air quality predictions can be most critical

Climatology

- Climatology is the study of average and extreme weather or air quality conditions at a given location.
- Climatology can help forecasters bound and guide their air quality predictions.

Climatology – Example

Average number of days per month with ozone in each AQI category for Sacramento, California



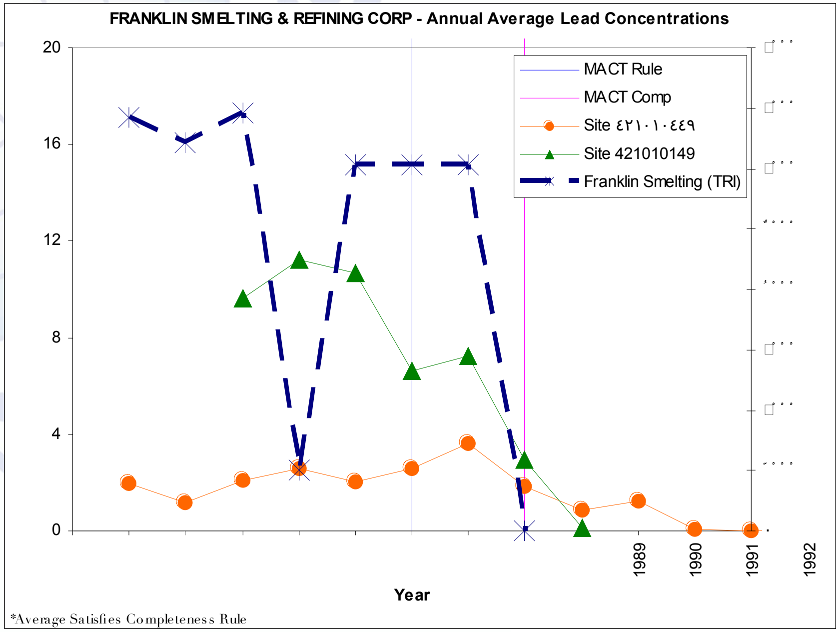
Developing Climatology

- Create a data set containing at least five years of recent pollutant data.
- Create tables or charts for forecast areas containing
  - All-time maximum pollutant concentrations (by month, by site)
  - Duration of high pollutant episodes (number of consecutive days, hours of high pollutant each day)
  - Average number of days with high pollutant levels by month and by week
  - Day-of-week distribution of high pollutant concentrations
  - Average and peak pollutant concentrations by holidays and non-holidays, weekends and weekdays

AREP
GAW

Developing Climatology

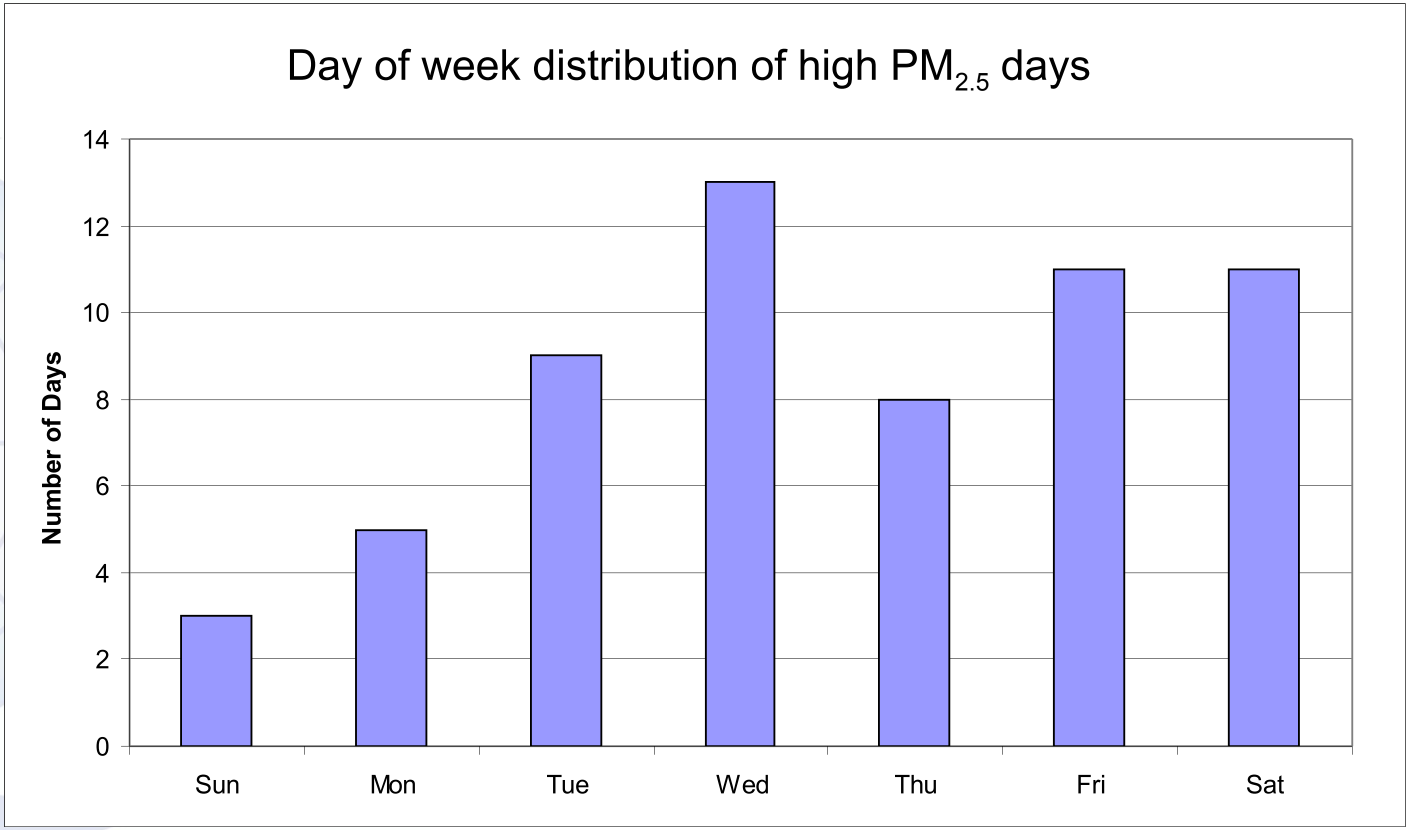
- Consider emissions changes
  - For example, fuel reformulation
  - It may be useful to divide the climate tables or charts into “before” and “after” periods for major emissions changes.



McCarthy et al., 2005

WMO
OMM

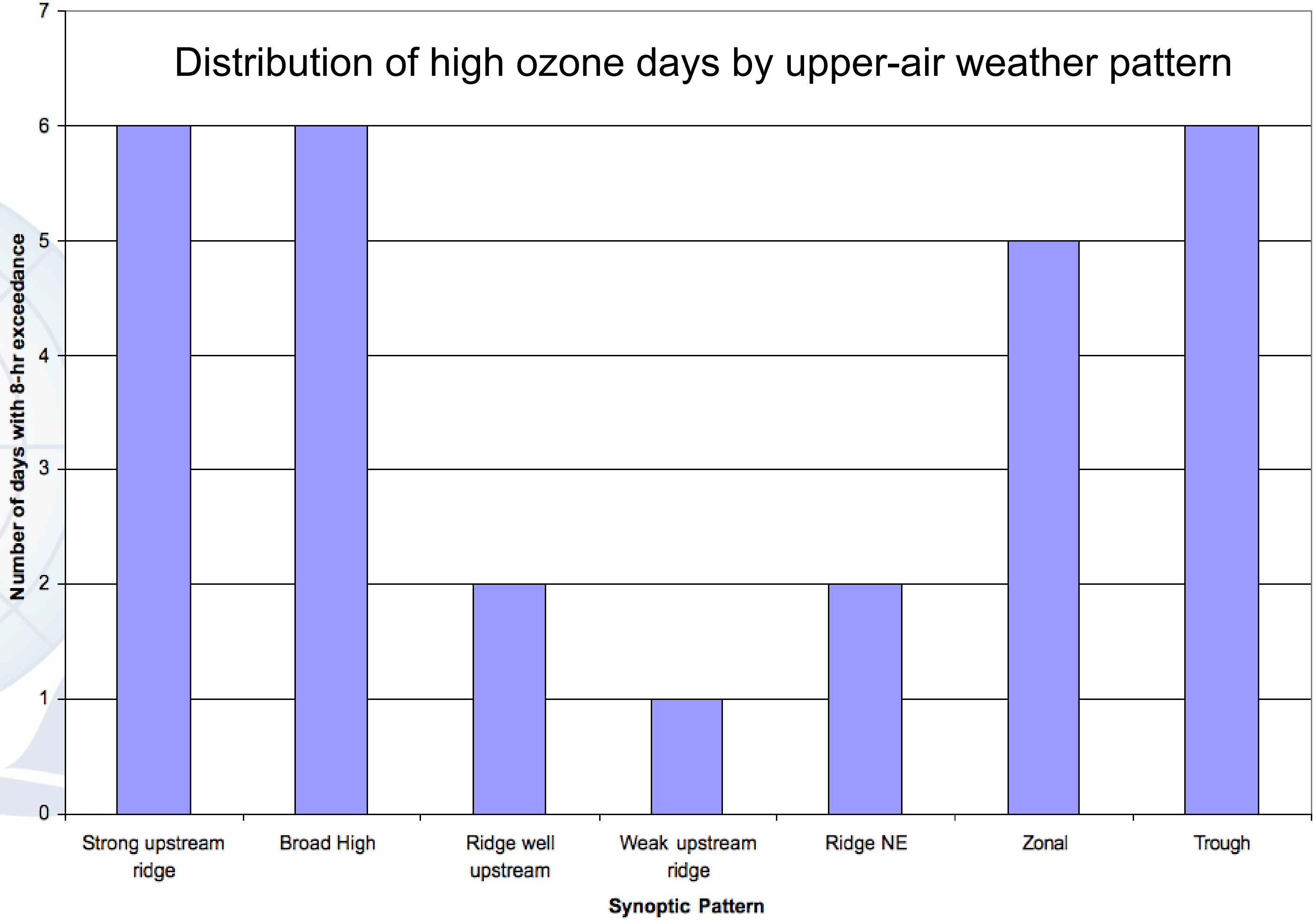
Climatology



- High pollution days occur most often on Wednesday
- High pollution days occur least often at the beginning of the week

AREP GAW

Climatology



WMO OMM

Climatology – Strengths

- Climatology
  - Bounds and guides an air quality forecast produced by other methods
  - Easy to develop

Climatology – Limitations

- Climatology
  - Is not a stand-alone forecasting method but a tool to complement other forecast methods
  - Does not account for abrupt changes in emissions patterns such as those associated with the use of reformulated fuel, a large change in population, forest fires, etc.
  - Requires enough data (years) to establish realistic trends

Criteria

- Uses threshold values (criteria) of meteorological or air quality variables to forecast pollutant concentrations
  - For example, if temperature > 27°C and wind < 2 m/s then ozone will be in the Unhealthy AQI category
- Sometimes called “rules of thumb”
- Commonly used in many forecasting programs as a primary forecasting method or combined with other methods
- Best suited to help forecast high pollution or low pollution events, or pollution in a particular air quality index category range rather than an exact concentration

Criteria – Example

Conditions needed for high pollution by month

Month	Daily Temp Max (above °C)	Daily Temp Range (above °C)	Daily Wind Speed (below m/s)	Wind Speed 15-21 UTC (below m/s)	Prior Day's Ozone 1-hr Max (above ppb)
Apr	26	11	4	3	70
May	29	11	4	5	70
Jun	29	11	3	5	70
Jul	33	11	3	4	70
Aug	33	11	3	4	70
Sep	31	10	3	4	75
Oct	31	10	3	3	75

Lambeth, 1998

To have a high pollution day in July

- maximum temperature must be at least 33° C,
- temperature difference between the morning low and afternoon high must be at least 11° C
- average daytime wind speed must be less than 3 m/s,
- afternoon wind speed must be less than 4 m/s, and
- the prior day's peak 1-hr ozone concentration must be at least 70 ppb.

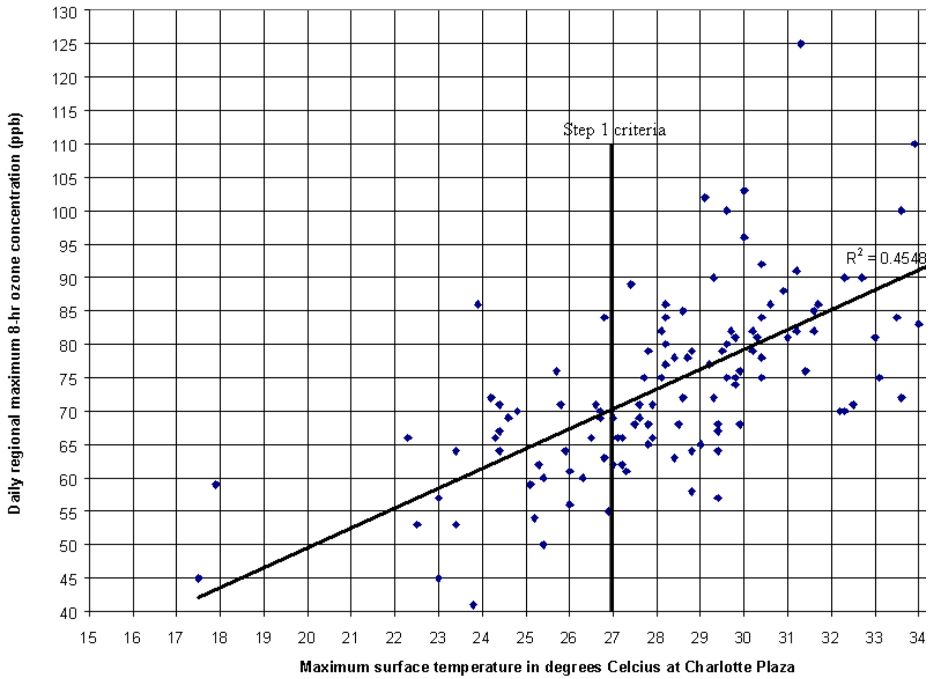
Developing Criteria (1 of 2)

- Determine the important physical and chemical processes that influence pollutant concentrations
- Select variables that represent the important processes
  - Useful variables may include maximum temperature, morning and afternoon wind speed, cloud cover, relative humidity, 500-mb height, 850-mb temperature, etc.
- Acquire at least three years of recent pollutant data and surface and upper-air meteorological data

AREP
GAW

Developing Criteria (2 of 2)

- Determine the threshold value for each parameter that distinguishes high and low pollutant concentrations.
- For example, create scatter plots of pollutant vs. weather parameters.
- Use an independent data set (i.e., a data set not used for development) to evaluate the selected criteria.



WMO
OMM

Criteria – Strengths

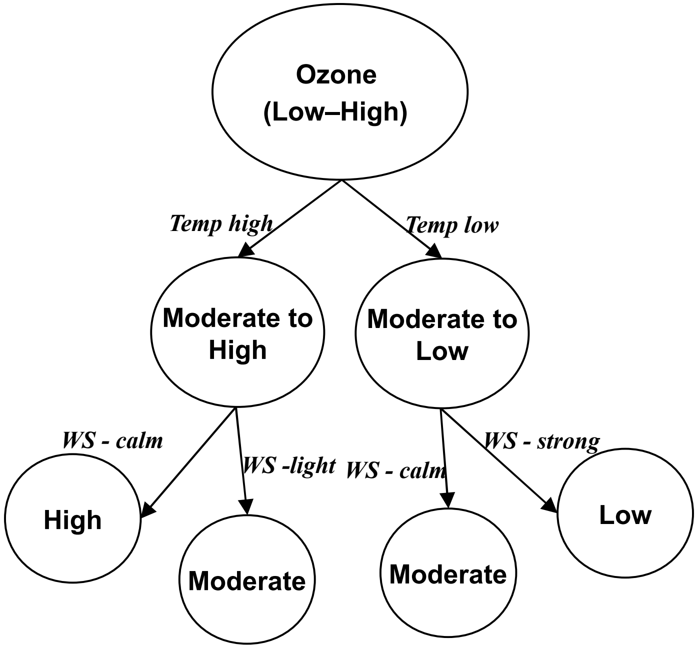
- Easy to operate and modify
- An objective method that alleviates potential human biases
- Complements other forecasting methods

Criteria – Limitations

- Selection of the variables and their associated thresholds is subjective.
- It is not well suited for predicting exact pollutant concentrations.

Classification and Regression Tree (CART)

- CART is a statistical procedure designed to classify data into dissimilar groups.
- Similar to criteria method; however, it is objectively developed.
- CART enables a forecaster to develop a decision tree to predict pollutant concentrations based on predictor variables (usually weather) that are well correlated with pollutant concentrations.



CART – How It Works (1 of 2)

The statistical software determines the predictor variables and the threshold cutoff values by

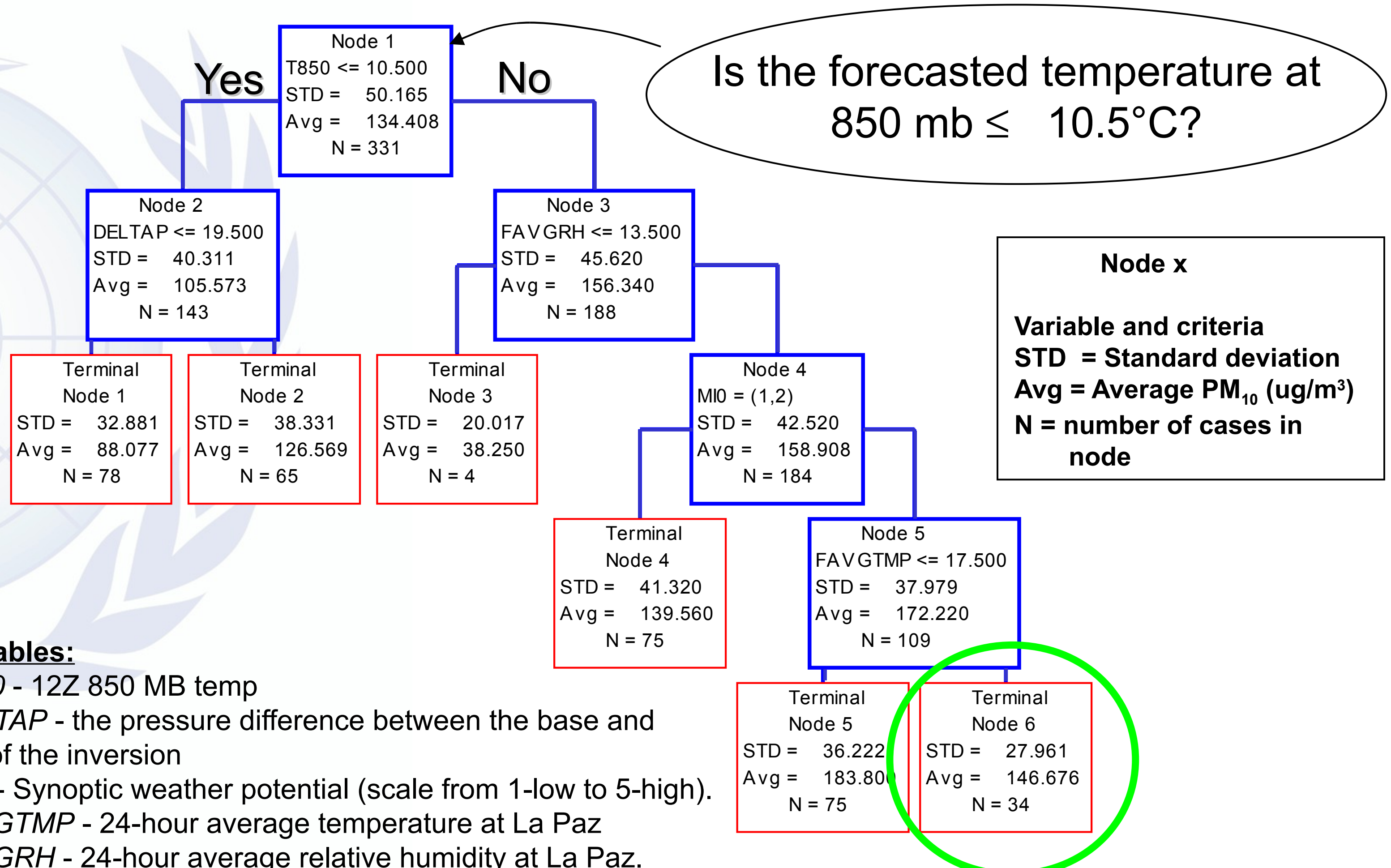
- Reading a large data set with many possible predictor variables
- Identifying the variables with the highest correlation with the pollutant
- Continuing the process of splitting the data set and growing the tree until the data in each group are sufficiently uniform

CART – How It Works (2 of 2)

- To forecast pollutant concentrations using CARTs
  - Step through the tree starting at the first split and determine which of the two groups the data point belongs in, based on the cut-point for that variable;
  - Continue through the tree in this manner until an end node is reached.
- The mean concentration shown in the end node is the forecasted concentration.
- Note, slight differences in the values of predicted variables can produce significant changes in predicted pollutant levels when the value is near the threshold.

CART – Example

CART classification PM_{10} in Santiago, Chile



Variables:

T850 - 12Z 850 MB temp
DELTAP - the pressure difference between the base and top of the inversion
MI0 - Synoptic weather potential (scale from 1-low to 5-high).
FAVGTMP - 24-hour average temperature at La Paz
FAVGRH - 24-hour average relative humidity at La Paz.

Cassmassi, 1999

Developing CART

- Determine the important processes that influence pollution.
- Select variables that properly represent the important processes.
- Create a multi-year data set of the selected variables.
  - Choose recent years that are representative of the current emission profile
  - Reserve a subset of the data for independent evaluation, but ensure it represents all conditions
  - Be sure variables are forecasted
- Use statistical software to create a decision tree.
- Evaluate the decision tree using the independent data set.

CART – Strengths

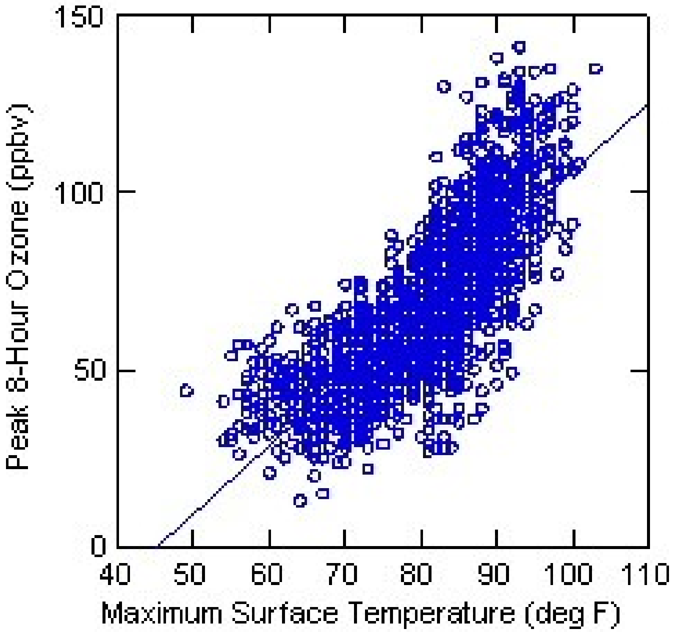
- Requires little expertise to operate on a daily basis; runs quickly.
- Complements other subjective forecasting methods.
- Allows differentiation between days with similar pollutant concentrations if the pollutant concentrations are a result of different processes. Since PM can form through multiple pathways, this advantage of CART can be particularly important to PM forecasting.

CART – Limitations

- Requires a modest amount of expertise and effort to develop.
- Slight changes in predicted variables may produce large changes in the predicted concentrations.
- CART may not predict pollutant concentrations during periods of unusual emissions patterns due to holidays or other events.
- CART criteria and statistical approaches may require periodic updates as emission sources and land use changes.

Regression Equations – How They Work (1 of 5)

- Regression equations are developed to describe the relationship between pollutant concentration and other predictor variables
- For linear regression, the common form is
  y = mx + b
- At right, maximum temperature (T_{max}) is a good predictor for peak ozone



$$[O_3] = 1.92 * T_{max} - 86.8$$

$$r = 0.77 \qquad r^2 = 0.59$$

Regression Equations – How They Work (2 of 5)

- More predictors can be added ("stepwise regression") so that the equation looks like this:

  $$y = m_1x_1 + m_2x_2 + m_3x_3 + \ldots\ldots m_nx_n + b$$

- Each predictor (x_n) has its own "weight" (m_n) and the combination may lead to better forecast accuracy.
- The mix of predictors varies from place to place.

Regression Equations – How They Work (3 of 5)

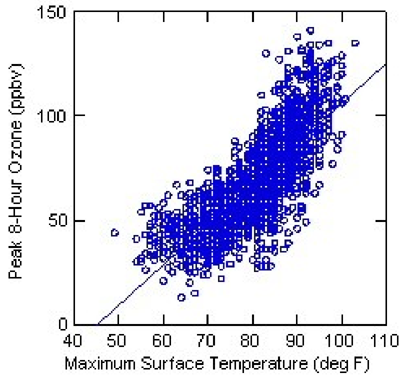
Ozone Regression Equation for Columbus, Ohio

$$8hrO_3 = \exp(2.421 + 0.024*Tmax + 0.003*Trange - 0.006*WS1to6 + 0.007*00ZV925 - 0.004*RHSfc00 - 0.002*00ZWS500)$$

Variable	Description
Tmax	Maximum temperature in ºF
Trange	Daily temperature range
WS1to6	Average wind speed from 1 p.m. to 6 p.m. in knots
00ZV925	V component of the 925-mb wind at 00Z
RHSfc00	Relative humidity at the surface at 00Z
00ZWS500	Wind speed at 500 mb at 00Z

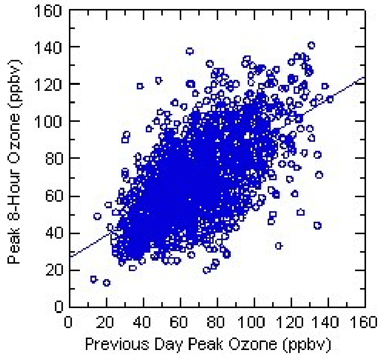
Regression Equations – How They Work (4 of 5)

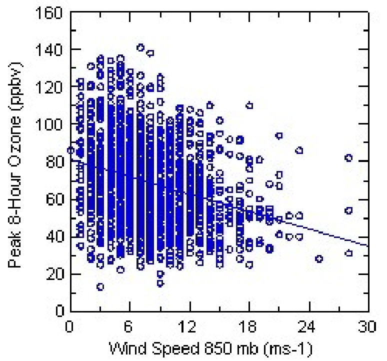
- The various predictors are not equally weighted, some are more important than others.
- It is essential to identify the strongest predictors and work hardest on getting those predictions right.



T_{max} vs. O_3

Previous day O_3 vs. O_3

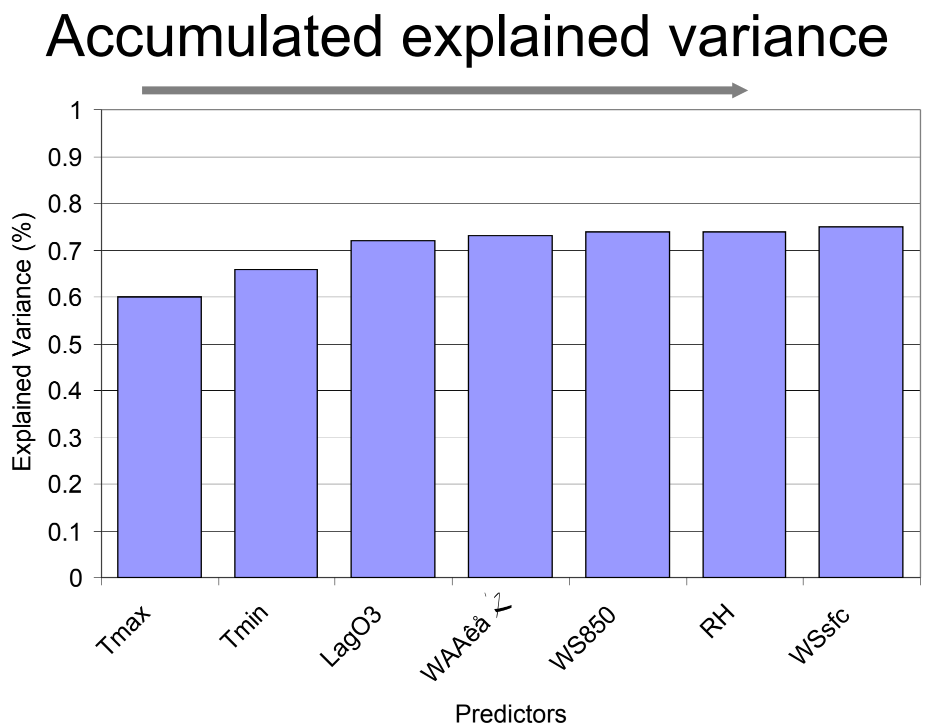




Wind Speed vs. O_3

Regression Equations – How They Work (5 of 5)

- In an example case, most of the variance in O_3 is explained by T_{max} (60%), with the additional predictors adding ~ 15%.
- Overall, 75% of the variance in observed O_3 is explained by the forecast model.
- Our job as forecasters is to fill in the additional 25% using other tools.



Developing Regression (1 of 2)

- Determine the important processes that influence pollutant concentrations.
- Select variables that represent the important processes that influence pollutant concentrations.
- Create a multi-year data set of the selected variables.
  - Choose recent years that are representative of the current emission profile.
  - Reserve a subset of the data for independent evaluation, but ensure it represents all conditions.
  - Be sure variables are forecasted.
- Use statistical software to calculate the coefficients and a constant for the regression equation.
- Perform an independent evaluation of the regression model.

Developing Regression (2 of 2)

- Using the natural log of pollutant concentrations as the predictand may improve performance.
- Do not to "over fit" the model by using too many prediction variables. An "over-fit" model will decrease the forecast accuracy. A reasonable number of variables to use is 5 to 10.
- Unique variables should be used to avoid redundancy and co-linearity.
- Stratifying the data set may improve regression performance.
  - Seasons
  - Weekend vs. weekday

Regression – Strengths

- It is well documented and widely used in a variety of disciplines.
- Software is widely available.
- It is an objective forecasting method that reduces potential biases arising from human subjectivity.
- It can properly weight relationships that are difficult to subjectively quantify.
- It can be used in combination with other forecasting methods, or it can be used as the primary method.

Regression – Limitations

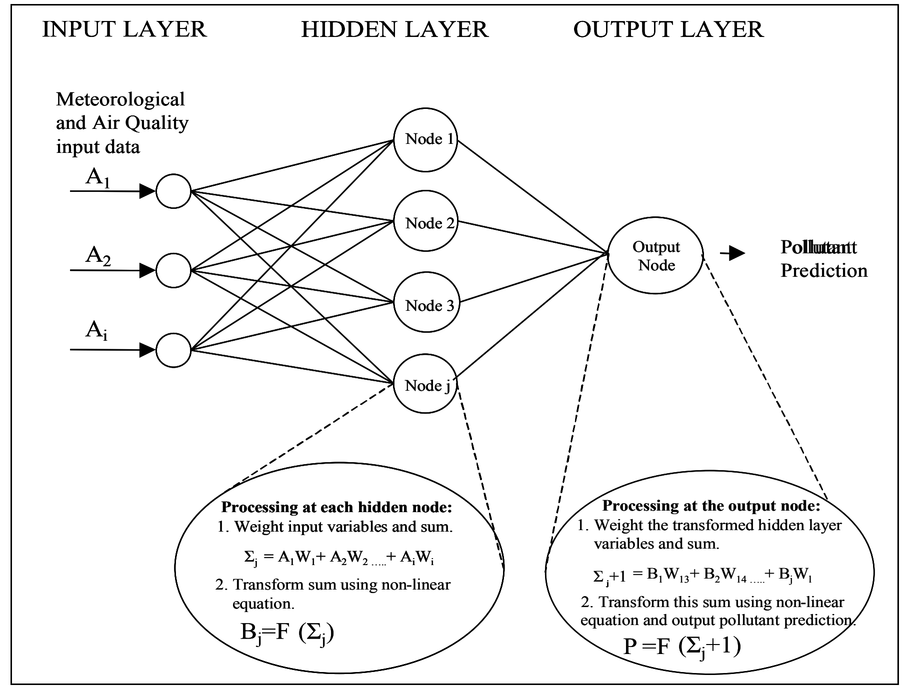
- Regression equations require a modest amount of expertise and effort to develop.
- Regression equations tend to predict the mean better than the tails (i.e., the highest pollutant concentrations) of the distribution. They will likely underpredict the high concentrations and overpredict the low concentrations.
- Regression criteria and statistical approaches may require periodic updates as emission sources and land use changes.
- Regression equations require 3-5 years of measurement data in the region of application, including many instances of air pollution events, to develop.

Neural Networks

- Artificial neural networks are computer algorithms designed to simulate the human brain in terms of pattern recognition.
- Artificial neural networks can be "trained" to identify patterns in complicated non-linear data.
- Because pollutant formation processes are complex, neural networks are well suited for forecasting.
- However, neural networks require about 50% more effort to develop than regression equations and provide only a modest improvement in forecast accuracy (Comrie, 1997).

Neural Networks – How It Works

- Neural networks use weights and functions to convert input variables into a prediction.
- A forecaster supplies the neural network with meteorological and air quality data.
- The software then weights each datum and sums these values with other weighted datum at each hidden node.
- The software then modifies the node data by a non-linear equation (transfer function).
- The modified data are weighted and summed as they pass to the output node.
- At the output node, the software modifies the summed data using another transfer function and then outputs a prediction.



Comrie, 1997

Developing Neural Networks

- Determine the important processes that influence pollutant concentrations.
- Select variables that represent the important processes.
- Create a multi-year data set of the selected variables.
  - Choose recent years that are representative of the current emission profile.
  - Reserve a subset of the data for independent evaluation, but ensure it represents all conditions.
  - Be sure variables are forecasted.
- Train the data using neural network software. See Gardner and Dorling (1998) for details.
- Test the trained network on a test data set to evaluate the performance. If the results are satisfactory, the network is ready to use for forecasting.

Neural Networks – Strengths

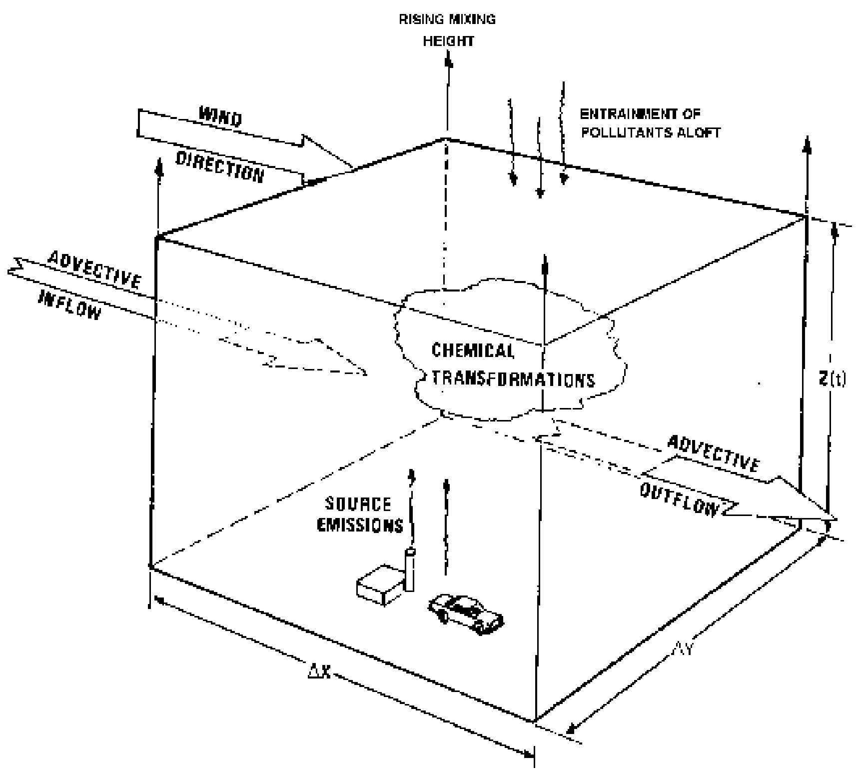
- Can weight relationships that are difficult to subjectively quantify
- Allows for non-linear relationships between variables
- Predicts extreme values more effectively than regression equations, provided that the network developmental set contains such outliers
- Once developed, a forecaster does not need specific expertise to operate it
- Can be used in combination with other forecasting methods, or it can be used as the primary forecasting method

Neural Networks – Limitations

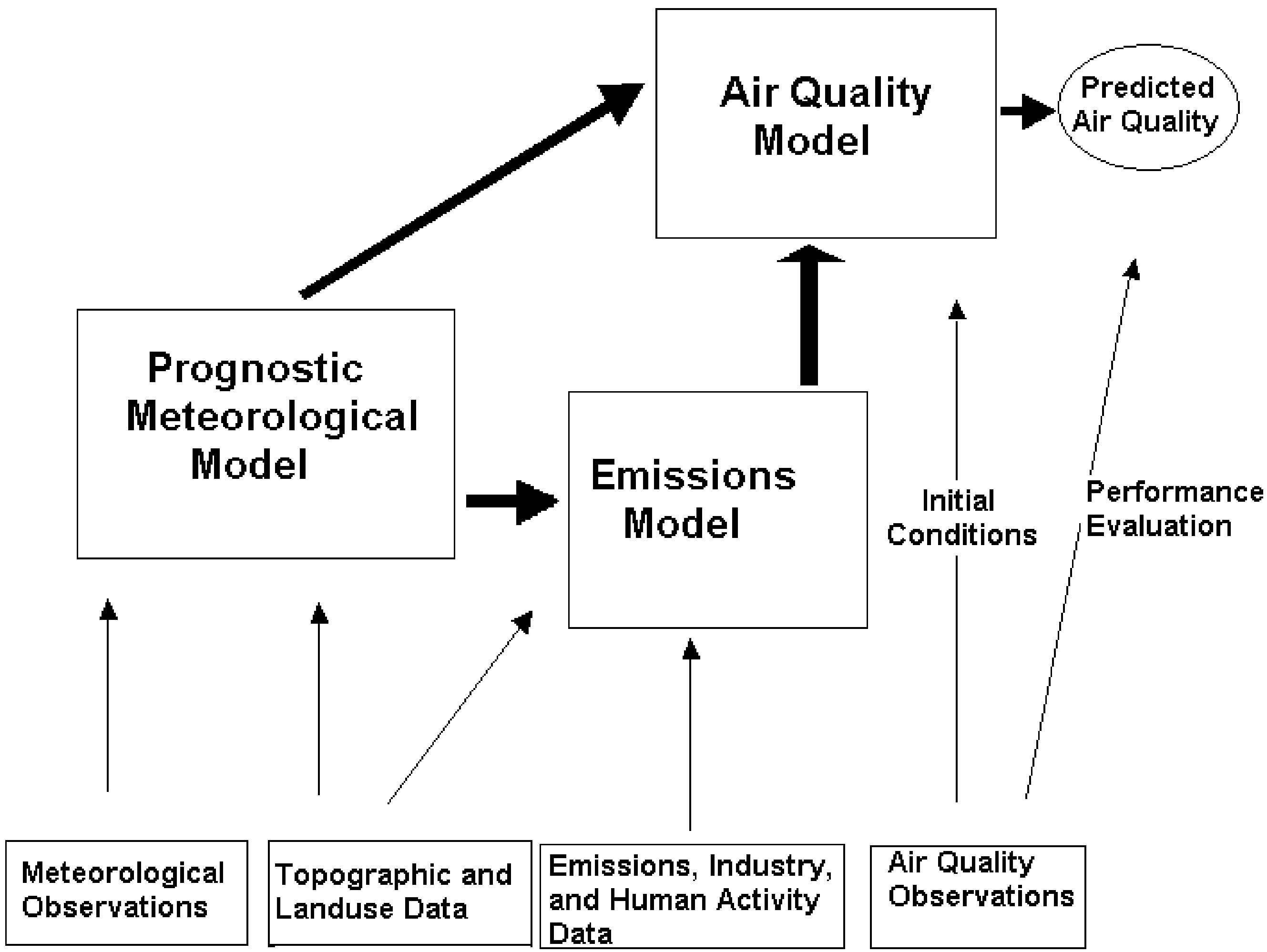
- Complex and not commonly understood; thus, the method can be inappropriately applied and difficult to develop
- Do not extrapolate data well; thus, extreme pollutant concentrations not included in the developmental data set will not be taken into consideration in the formulation of the neural network prediction
- Require 3-5 years of measurement data in the region of application, including many instances of air pollution events, to develop.

Numerical Modeling

- Mathematically represents the important processes that affect pollution
- Requires a system of models to simulate the emission, transport, diffusion, transformation, and removal of air pollution
  - Meteorological forecast models
  - Emissions models
  - Air quality models



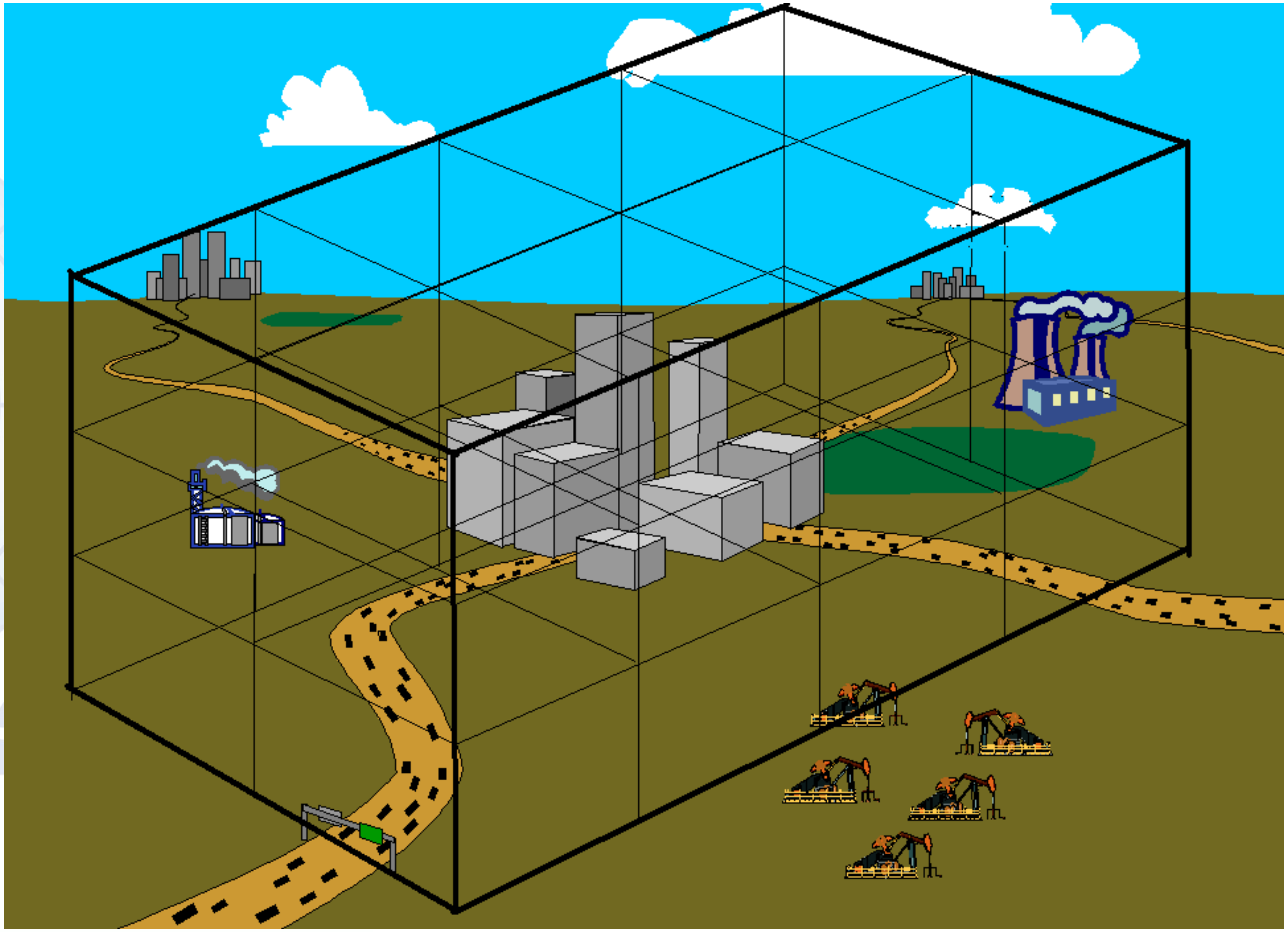
Numerical Modeling – How It Works



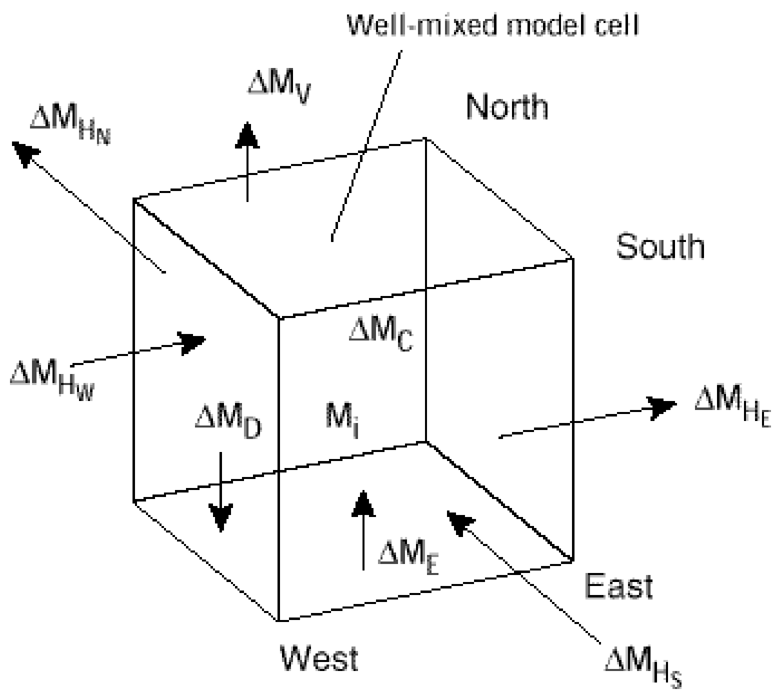
Processes Treated in Grid Models

- *Emissions*
  - Surface emitted sources (on-road and non-road mobile, area, low-level point, biogenic, fires)
  - Point sources (electrical generation, industrial, other, fires)
- *Advection (Transport)*
- *Dispersion (Diffusion)*
- *Chemical Transformation*
  - VOC and NO_x chemistry, radical cycle
  - For PM aerosol thermodynamics and aqueous-phase chemistry
- *Deposition*
  - Dry deposition (gas and particles)
  - Wet deposition (rain out and wash out, gas and particles)

Photochemical Grid Model Concept

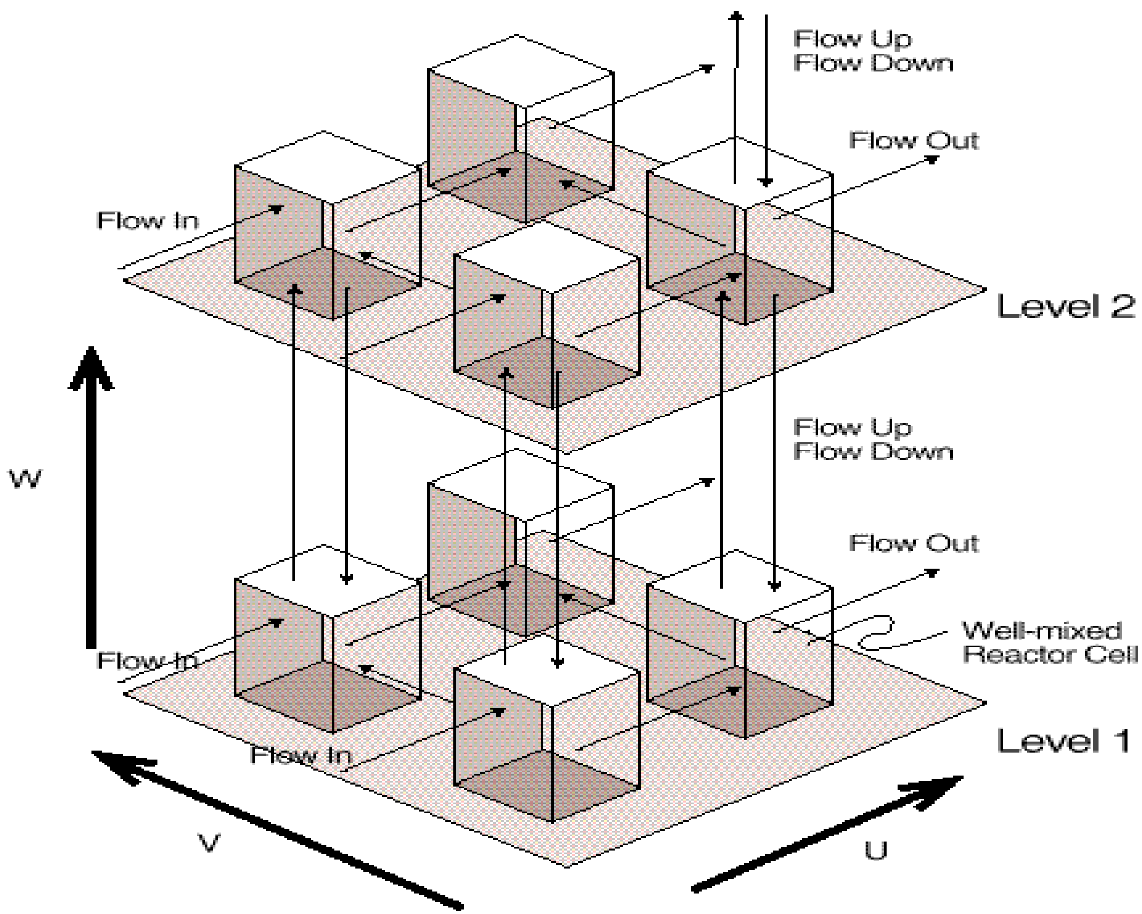


Eulerian Grid Cell Processes

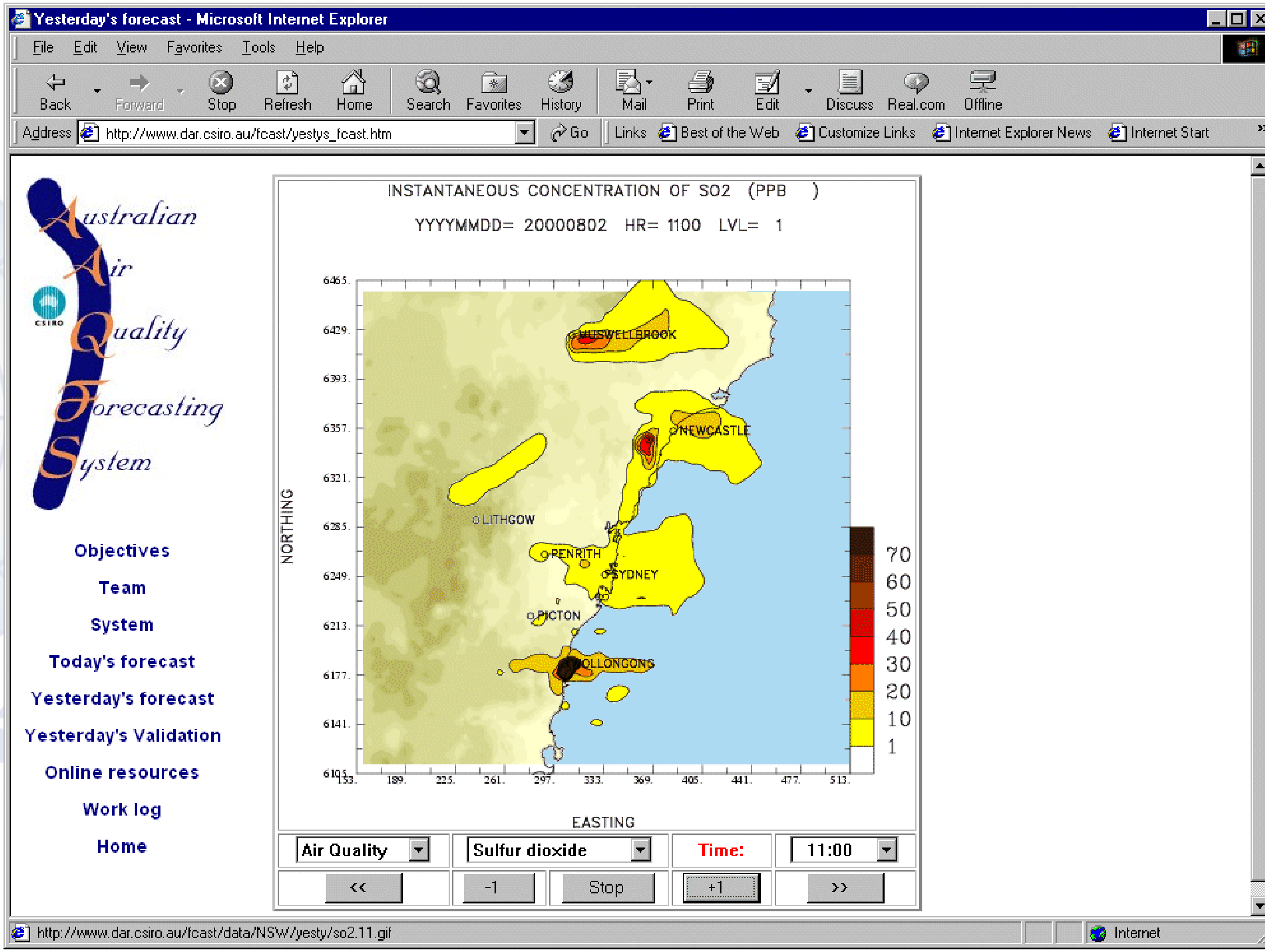


H is horizontal transport
V is vertical transport
E is emissions
D is surface deposition
C is chemical transformations
i is initial
f is final
M is species mixing ratios

Coupling Between Grid Cells



Numerical Modeling – Example



Developing a Numerical Model

- Design and plan the system
- Identify and allocate the resources
- Acquire required geophysical data
- Implement the data acquisition and processing tools, component models (emissions, meteorological, and air quality), and analysis programs.
- Develop the emission inventory
- Test the operation of all data acquisition programs, preprocessor programs, component models, and analysis programs as a system
- Integrate data acquisition and processing tools, component models, and analysis programs into an operational system
- Test, evaluate, and improve the integrated system

Developing a Numerical Model (1 of 7)

Design and plan the system

- Decide on which pollutants to forecast.
- Define modeling domains considering geography and emissions sources.
- Select component models considering forecast pollutants, domains, component model compatibility, availability of interface programs, and available resources.
- Determine hardware and software requirements.
- Identify sources of meteorological, emissions, and air quality data.
- Prepare a detailed plan for acquiring and integrating data acquisition, modeling, and analysis software.
- Plan for continuous real-time evaluation of the modeling system.

Developing a Numerical Model (2 of 7)

Identify and allocate the resources

- Staff for system implementation and operations
- Computing and storage consistent with the selection of domains and models
- Communications for data transfer into and out of the modeling system

Acquire required geophysical data

- Topographical data
- Land use data

Developing a Numerical Model (3 of 7)

Implement the data acquisition and processing tools, component models (emissions, meteorological, and air quality), and analysis programs.

- Implement each program individually.
- Use standard test cases to verify correct implementation.

Developing a Numerical Model (4 of 7)

Develop the emission inventory

- Acquire needed emission inventory related data.
- Review the emissions data for accuracy.
- Be sure that the emission inventory includes the most recent emissions data available.
- Update the base emission inventory annually.

Developing a Numerical Model (5 of 7)

Test and Evaluate

- Test the operation of all data acquisition programs, preprocessor programs, component models, and analysis programs as a system.
- Review the prognostic meteorological forecast data for accuracy over several weeks under various weather patterns.
- Run the combined meteorological/emissions/air quality modeling system in a prognostic mode using a variety of meteorological and air quality conditions.
- Evaluate the performance of the modeling system by comparing it with observations.
- Refine the model application procedures (i.e., the methods of selecting boundary conditions or initial concentration fields, the number of spin-up days, the grid boundaries, etc.) to improve performance.

Developing a Numerical Model (6 of 7)

Integrate data acquisition and processing tools, component models, and analysis programs into an operational system

- Implement automated processes for data acquisition, the daily data exchange from the prognostic meteorological model and the emissions model to the 3-D air quality model and analysis programs, and forecast product production.
- Implement automated processes by using scripting and scheduling tools.
- Verify that the forecast products reflect the actual model predictions.

Developing a Numerical Model (7 of 7)

Test, evaluate, and improve the integrated system

- Run the model in real-time test mode for an extended period. Compare output to observed data and note when there are model failures.
- After obtaining satisfactory results on a consistent basis, use the modeling system to forecast pollutant concentrations.
- Document the modeling system.
- Continuously evaluate the system's performance by comparing observations and predictions.
- Implement improvements as needed based on performance evaluations and new information.

Numerical Modeling – Strengths

- They are phenomenological based, simulating the physical and chemical processes that result in the formation and destruction of air pollutants.
- They can forecast for a large geographic area.
- They can predict air pollution in areas where there are no air quality measurements.
- The model forecasts can be presented as maps of air quality to show how predicted air quality varies over a region hour by hour.
- The models can be used to further understand the processes that control air pollution in a specific area. For example, they can be used to assess the importance of local emissions sources or long-range transport.

Numerical Modeling – Limitations

- Inaccuracies in the prognostic model forecasts of wind speeds, wind directions, extent of vertical mixing, and solar insulation may limit 3-D air quality model performance.
- Emission inventories used in current models are often out of date and based on uncertain emission factors and activity levels.
- Site-by-site ozone concentrations predicted by 3-D air quality forecast models may not be accurate due to small-scale weather and emission features that are not captured in the model.
- Substantial staff and computer resources are needed to establish a scientifically sound and automated air quality forecast system based on a 3-D air quality model.

AREP
GAW

Australian Air Quality Forecasting System

Peter Manins

CSIRO Marine and Atmospheric Research
Australia

WMO GURME SAG member

Natural Heritage Trust

OMM

Demonstration Project

EPA ENVIRONMENT PROTECTION AUTHORITY NSW

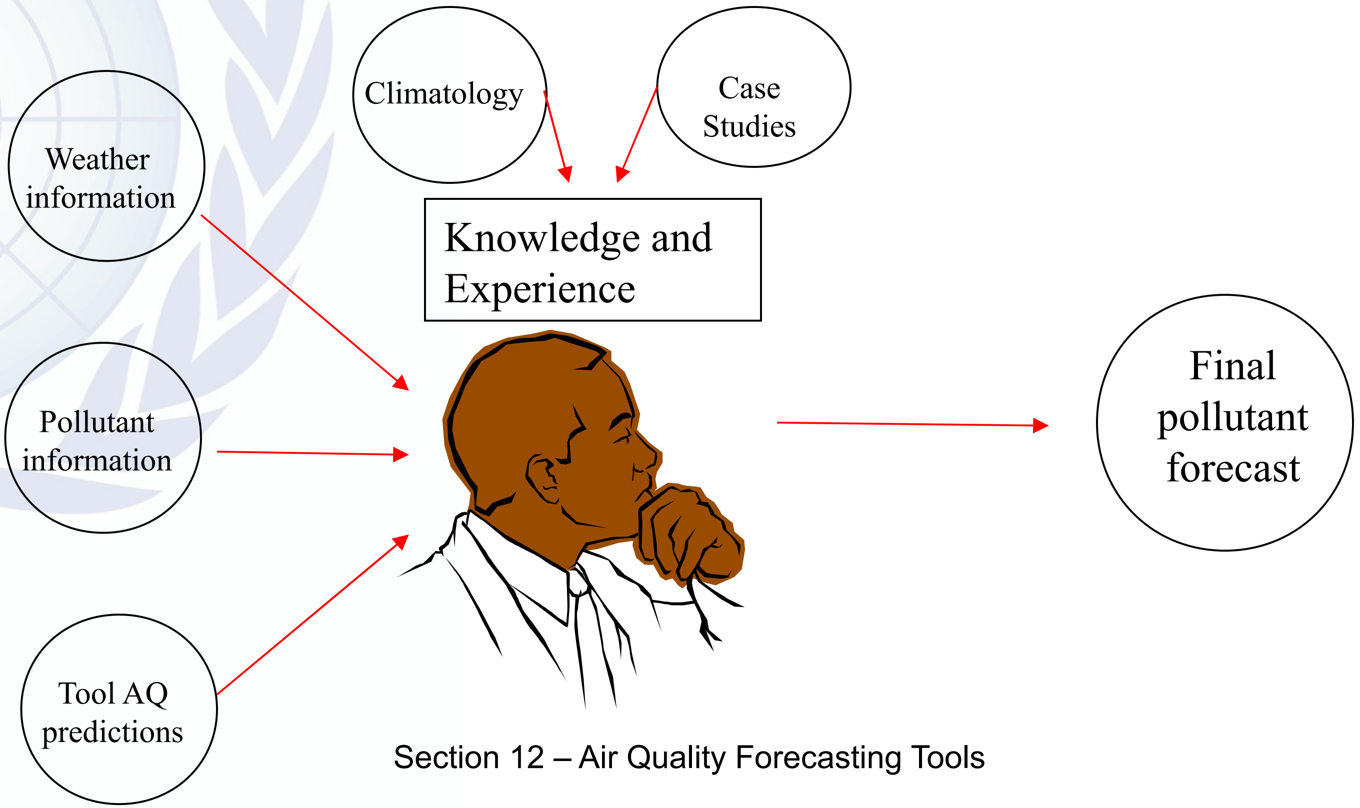
CSIRO

BMRC Bureau of Meteorology Research Centre

Environment Australia

Phenomenological – How It Works

- Relies on forecaster experience and capabilities
- Forecaster needs good understanding of the processes that influence pollution such as the synoptic, regional, and local meteorological conditions, plus air quality characteristics in the forecast area.
- Forecaster synthesizes the information by analyzing observed and forecasted weather charts, satellite information, air quality observations, and pollutant predictions from other methods to develop a forecast.



Phenomenological/Intuition

- Involves analyzing and conceptually processing air quality and meteorological information to formulate an air quality prediction.
- Can be used alone or with other forecasting methods such as regression or criteria.
- Is heavily based on the experience provided by a meteorologist or air quality scientist who understands the phenomena that influence pollution.
- This method balances some of the limitations of objective prediction methods (i.e., criteria, regression, CART, and neural networks).

AREP
GAW

Phenomenological – Example

Knowledge can be documented as forecast rules

Boston	PM$_{2.5}$ Forecasting Guidelines for November - March			
	0-15 µg/m^3	16-40 µg/m^3	41-65 µg/m^3	66-150 µg/m^3
	Good	Moderate	Unhealthy for Sensitive Groups (USG)	Unhealthy
Upper-air Pattern	Strong 500-mb low-pressure trough encompassing the Boston area or 500-mb zonal flow encompassing the Boston area or 500-mb cut-off low-pressure system encompassing eastern Massachusetts	Well-defined 500-m[illegible] encompassing the B[illegible] or 500-mb zonal flow encompassing the Boston area or 500-mb short-wave trough passing over or just south of the Boston area following a Moderate day [illegible] 500-mb cut-off low-pressure system located in the Mid-Atlantic States, south of Boston	500-mb ridge [illegible] the Boston area or 500-mb zonal flow encompassing the Boston area	No Unhealthy days occurred during the last 3 "winters"
Inversion Strength and Duration	Weak or no inversion below 900 mb at 12Z (0700 EST). Inversion may or may not be present above 900 mb	[illegible]rface in[illegible] (0700 EST) that ma[illegible] day	[illegible]face inversion at 12Z (0700 EST) that is not likely to break during the day	
Surface Wind	Strong westerly-northerly winds (> 10 kts) for the majority of the day	Light to moderate surface winds (0-10 kts) from any direction for the majority of the day or Moderate to strong south-southwesterly surface winds (5-20 kts) for the ma[illegible] causing transport from the south	Light surface winds (0-4 kts) from any direction for the majority of the day or Moderate to strong south-southwesterly surface winds (5-20 kts) for the majority of the day causing transport from the [illegible]	
Other	Zonal flow or a moderate ridge at 500 mb, light surface winds, and a previous-day AQI value in the low "Good" category or Well-defined surface front passed through the area on the last day +	Stationary front located in northeastern Massa[illegible] Boston + Consider timing of frontal passage – early high concentrations can outweigh the post fr[illegible] concentrations Local convergence or divergence aloft can negate effects of weak upper-level trough or ridge	Stationary front located in northeastern Massachusetts, north of Boston + Only 3 USG events occurred during the last 3 "winters"	
Previous-day AQI Category	In general the air quality conditions increase and decrease by only one AQI category per day			

+ indicates meteorological variables that tend to overwhelm other factors when present.

Boston	
	0-15 µg/m^3
	Good
Upper-air Pattern	Strong 500-mb low-pressure trough encompassing the Boston area or 500-mb zonal flow encompassing the Boston area or 500-mb cut-off low-pressure system encompassing eastern Massachusetts
Inversion Strength and Duration	Weak or no inversion below 900 mb at 12Z (0700 EST). Inversion may or may not be present above 900 mb
Surface Wind	Strong westerly-northerly winds (> 10 kts) for the majority of the day
Other	Zonal flow or a moderate ridge at 500 mb, light surface winds, and a previous-day AQI value in the low "Good" category or Well-defined surface front passed through the area on the last day +
Previous-day AQI Category	In general the air quality conditi

Phenomenological – Forecast Worksheets

Example forecast worksheets for $PM_{2.5}$

	$PM_{2.5}$ (ug/m^3)	500 mb pattern	Surface pattern	Winds	Inversion/ Mixing	Carryover	Clouds/ fog	Transport/ Recirculation
Yesterday	46	Ridge	High pressure	Light and variable	Strong inversion	Yes	Mix	No
Today	70	Weak ridge	High pressure	Light and variable	Strong inversion	Yes	Mix	No
Tomorrow	35	Weak trough	Cold front/ trough	Moderate west-northwest	No inversion	Yes	Mix	No

$PM_{2.5}$ Forecast (μg/m³)	Phenomenological	Objective Tool	Final
Today	70	42	65
Tomorrow	35	18	25

Developing Phenomenological

- Key step is acquiring an understanding of the important physical and chemical processes that influence pollution in your area.
  - Literature reviews
  - Historical case studies of air quality events
  - Climatological analysis
- Although much knowledge can be gleaned from these sources, the greatest benefit to the method is gained through forecasting experience.

Phenomenological – Strengths

- Allows for easy integration of new data sources
- Allows for the integration and selective processing of large amounts of data in a relatively short period of time
- Can be immediately adjusted as new truths are learned about the processes that influence ozone or $PM_{2.5}$
- Allows for the effect of unusual emissions patterns associated with holidays and other events to easily be taken into account
- Is better for extreme or rare events. Generally, objective methods such as regression or neural networks do not capture extreme or rare events
- Is a good complement to other more objective forecasting methods because it tempers their results with common sense and experience

Phenomenological – Limitations

- Requires a high level of expertise.
  - The forecaster needs to have a strong understanding of the processes that influence pollution.
  - The forecaster needs to apply this understanding in both the developmental and operational processes of this method.
- Forecaster bias is likely to occur. Using an objective method as a complement to this method can alleviate these biases.

Selecting Predictor Variables (1 of 3)

- Many methods require predictor variables.
  - Meteorological
  - Air Quality
- Before selecting particular variables it is important to understand the phenomena that affect pollutant concentrations in your region.
- The variables selected should capture the important phenomena that affect pollutant concentrations in the region.

Selecting Predictor Variables (2 of 3)

- Select observed and forecasted variables. Predictor variables can consist of observed variables (e.g., yesterday’s ozone or $PM_{2.5}$ concentration) and forecasted variables (e.g., tomorrow’s maximum temperature).
- Make sure that predictor variables are easily obtainable from reliable source(s) and can be forecast.
- Consider uncertainty in measurements, particularly measurements of PM.

Selecting Predictor Variables (3 of 3)

- Begin with as many as 50 to 100 predictor variables.
- Use statistical analysis techniques to identify the most important variables.
  - Cluster analysis is used to partition data into similar and dissimilar subsets. Unique (i.e., dissimilar) variables should be used to avoid redundancy.
  - Correlation analysis is used to evaluate the relationship between the predictand (i.e., pollutant levels) and various predictor variables.
  - Step-wise regression is an automatic procedure that allows the statistical software (SAS, Statgraphics, Systat, etc.) to select the most important variables and generate the best regression equation.
  - Human selection is another means of selecting the most important predictor variables.

Common Ozone Predictor Variables

Variable	Usefulness	Condition for High Ozone
Maximum temperature	Highly correlated with ozone and ozone formation	High
Morning wind speed	Associated with dispersion and dilution of ozone precursor pollutants	Low
Afternoon wind speed	Associated with transport of ozone	-
Cloud cover	Controls solar radiation, which influences photochemistry	Few
Relative humidity	Surrogate for cloud cover	Low
500-mb height	Indicator of the synoptic-scale weather pattern	High
850-mb temperature	Surrogate for vertical mixing	High
Pressure gradients	Causes winds/ventilation	Low
Length of day	Amount of solar radiation	Longer
Day of week	Emissions differences	-
Morning NO_x concentration	Ozone precursor levels	High
Previous day's peak ozone concentration	Persistence, carry-over	High
Aloft wind speed and direction	Transport from upwind region	-

AREP GAW

Common $PM_{2.5}$ Predictor Variables

Variable	Usefulness	Condition for High $PM_{2.5}$
500-mb height	Indicator of the synoptic-scale weather pattern	High
Surface wind speed	Associated with dispersion and dilution of pollutants	Low
Surface wind direction	Associated with transport of pollutants	-
Pressure gradient	Causes wind/ventilation	Low
Previous day's peak $PM_{2.5}$ concentration	Persistence, carry-over	High
850-mb temperature	Surrogate for vertical mixing	High
Precipitation	Associated with clean-out	None or light
Relative Humidity	Affects secondary reactions	High
Holiday	Additional emissions	-
Day of week	Emissions differences	-

WMO OMM

Summary

- Wide range of forecast tools
- Each type has advantages and disadvantages
- More tools result in better forecasts
- Consensus forecasting can produce better results