Statistical forecasting techniques to describe the surface winds in Sydney Harbour

Gregory J. Connor
Bureau of Meteorology, Townsville, Queensland, Australia

Elly Spark
Bureau of Meteorology, Sydney, New South Wales, Australia

and

William T.M. Dunsmuir
Department of Statistics, University of New South Wales, Sydney, Australia

(Manuscript received October 2002; revised April 2003)

Sydney Harbour was selected as the venue for the sailing events for the 2000 Olympic and Paralympic games. A suite of statistically based prediction methods were derived specifically for these events and provided forecasts for surface wind speed, wind direction, wind gust, and sea-breeze occurrence. The statistical models derived relationships from historical data using linear, logistic and non-linear regression, and also classification and regression trees. A feature of the statistical prediction models was the use of predictors that were selected for their meteorological relevance to the low-level wind structure. Emphasis was placed upon predictors that were influential in defining the sea-breeze circulation in previous studies.

During the 1999 Olympic Test Events (from 16-27 September 1999), in a time-lagged trial using observed gradient wind data, a regression based statistical tool was verified with root mean square error values of 40 degrees for direction and 2.9 kn for speed (1 kn = 0.51 m s⁻¹). During a period that encompassed the Olympic and Paralympic Games, under real time operational conditions, a synoptically stratified Model Output Statistics (MOS) technique displayed superior skill over conventional MOS methods.

Introduction

Sydney Harbour, and the immediate offshore area, was chosen as the venue for sailing events for the 2000 Olympic and Paralympic Games. The sea-breeze is an intrinsic part of the Sydney Basin climatology. It is well established that the sea-breeze behaves similarly to a density current driven by a thermal contrast and subsequent density difference between land and sea (Simpson 1995; Atkins and Wakimoto 1997). The temperature contrast between land and sea has previously been used to quantify the onshore daytime wind component (Connor 1997).
Zhong and Takle (1992) suggest that this thermal contrast is more influential in situations of weak synoptic forcing. Numerical modelling studies such as Estoque (1962), and more recently, Arritt (1993) and Zhong and Takle (1993), have explored the link between the sea-breeze circulation and the prevailing synoptic flow. These studies discuss distinctive characteristics of the sea-breeze under different synoptic regimes. Observational studies (Wakimoto and Atkins 1994; Atkins and Wakimoto 1997) support this distinction in terms of onshore, parallel to shore, and offshore synoptic-scale flow. Their studies note that, provided the offshore flow is not too great, a more pronounced sea-breeze results under offshore synoptic flow conditions. They report a greater definition of the sea-breeze front and more constant propagation speeds in an offshore flow, when compared to onshore, or parallel to shore, conditions. Houghton (1992) discusses sea-breeze properties by defining quadrants, dependent on coastline orientation and the prevailing gradient wind.

The relationship between the surface wind and the gradient wind is given prominence in this paper and has provided the basis for the development of a range of statistical products for forecasting surface winds on Sydney Harbour. A number of forecasting tools are derived specifically for hourly wind forecasting for events at the Olympic and Paralympic Games. These forecasting tools share a broad methodology; statistical associations are determined between a historical dataset of observed surface winds and a temporally matched dataset of meteorological parameters. Both Perfect Prog (Klein et al. 1959) and Model Output Statistics (MOS) (Glahn and Lowry 1972) techniques are used with observed upper wind data, conventional numerical weather prognosis (NWP) parameters, and a range of other predictors that have been associated with coastal winds in previous studies (Connor 1997; Kim 1999). Most of the statistical forecasting tools rely on regression (linear and non-linear) to provide statistical links between the wind predictand and the historical datasets. Classification and Regression Trees (CART) (Breiman et al. 1984), a type of recursive partitioning model, is also employed in one of the forecasting tools. The CART technique has been successfully used to develop practical forecasting applications (Burrows 1991; Burrows et al. 1995; Carter and Elsner 1997; Connor and Woodcock 2000).

Pattern matching, or stratification of data on a synoptic basis, from a historical dataset has previously been used in the Australian context for rainfall and temperature forecasting (e.g., Dahni and Stern 1995; Connor and Woodcock 2000; Stern 1999). Some of the statistical techniques discussed use the gradient wind to partition data prior to the development of the statistical models. It is the aim of this process to define a pattern of synoptic flow and then capture predictors that are pertinent to this synoptic wind regime.

A description of the geographical setting of Sydney Harbour and a brief climatology of the September/October period is initially provided, followed by an explanation of the derivation of the statistical wind forecasting tools. The last section presents a summary and a discussion on the results.

Sydney Harbour

Geographical setting and climatology for September and October

Sydney Harbour is a region of complex shoreline and topography (Fig. 1), and although the harbour foreshores only rise to about 100 m, the surface winds are influenced greatly by the channelling effect of the harbour. The effect of the harbour on the wind flow is dependent on the atmospheric stability and is most obvious when the synoptic flows are relatively weak. Further insight can be gained through examining a 24-hour period that incorporates the evolution of a typical sea-breeze, where the harbour orography contributes to the complexity of the sea-breeze (Banta 1995).

Figure 2 shows the progression of the winds in the harbour region at 2-hour intervals up until 4 pm (1600 Eastern Standard Time (EST)), and then, at intervals of four hours. Light westerly winds or 'drainage flow' in the early morning appear to follow the harbour orientation. By 10 am, the winds stall as a sea-breeze becomes established in the outer harbour, first with very light winds, but then gradually strengthening to a

![Topographical map of Sydney Harbour indicating the sources of wind data. The areas marked A to F show the regions within which each sailing course was conducted.](image)

---

Fig. 1 Topographical map of Sydney Harbour indicating the sources of wind data. The areas marked A to F show the regions within which each sailing course was conducted.
Fig. 2 A case study of the evolution of the Sydney Harbour winds for a 24-hour period, 19-20 October 1997. The times shown are EST. The stations displayed are Fort Denison, Clarke Island, Shark Island, Wedding Cake, Middle Head, Bombo, Cannae Pt, South Head and North Head. The wind barbs show wind speed (kn) and direction and the numbers represent maximum gusts (kn).
Fig. 2  Continued.
Fig. 3  A time series displaying a vertical cross-section of the winds at Sydney Airport, separated into onshore and offshore components, for the period 13-16 September 1999. The onshore components were compiled from pilot balloon data at 1000 ft intervals. The winds at 0000, 1200, 1800 EST were interpolated from adjacent data. Onshore components are rendered as cool colours and offshore components as warm colours. The average gradient winds at 0900 EST for each day are also shown to give an indication of the prevailing synoptic flow. The sea-breeze on 13, 14 and 15 September can be clearly seen as the blue onshore current near 1500 EST. A return (offshore) flow is apparent on 13 and 14 September, centred at around 6000 ft.

maximum at around 4 pm. The winds then back and gradually ease and during the late afternoon with the sea-breeze direction tending to follow the harbour orientation. By 8 pm the winds have backed to the north, and by midnight the westerly flow is again evident. On the second day, a cold front moves in from the south, at about 4 am, and a southerly flow is established.

The sea-breeze of the Sydney Basin

Although the sea-breeze is not observed on the inner harbour on all occasions, Dunsmuir et al. (2003) record its presence on approximately 40 per cent of days during September. The presence of this thermal circulation is usually significant to marine users as it is indicated by a significant wind shift at sea level. In previous studies focusing on the Sydney Basin sea-breeze, Kim (1999) and Dunsmuir et al. (2003), expended considerable effort on the determination of sea-breeze occurrence – due to its dramatic effect on the local surface wind. The Sydney Harbour sea-breeze is not a homogeneous entity, there is great variation in the structure and evolution of this density current – depending on the synoptic setting. Daily variation in the strength and depth of the sea-breeze for the period 13-16 September 1999 is illustrated by the vertical atmospheric cross-section in Fig. 3.

The cross-section is constructed from upper wind observations taken over a period of almost four days and under varying gradient wind conditions. During the relatively light gradient wind days of 13 and 14 September 1999, the sea-breeze and associated return flow can be clearly discerned. The early morning land breeze/drainage flow is also evident at around 0600 EST at Sydney Airport (approximately 10 km to the south of Sydney City). The time series displayed in Fig. 3 indicates the sea-breeze peaks at around 1500 EST and is variable in depth, ranging from over 3000 ft on 13 September to just above 1000 ft two days later. A more general indication of the structure of the sea-breeze at Sydney Airport in September can be
seen in Fig. 4, where, a composite vertical cross-section is constructed for the 13 days when a significant wind shift, that could be attributed to the sea-breeze, did occur during the 28-day period between 3 and 30 September 2000. The composite clearly shows an onshore current centred near 1500 EST and with a depth of around 1500 ft. The return flow is apparent above 2000 ft, with a maximum between 5000 and 8000 ft, where the diurnal variation in the offshore component is clearly evident.

**Statistical techniques**

The relationship between the gradient wind and the wind at the surface is fundamental to the suite of statistical methods derived in this study. All of the surface wind forecasting methodologies described below require a forecast gradient wind. The rationale behind considering the synoptic setting at gradient level is explored in Fig. 5 (upper panel), where hourly surface winds from the Offshore Reference Station (ORS), comprising of August, September and October observations for the years 1991-99, was matched with the 2300 UTC (0900 EST) gradient wind (900 hPa) data from Sydney Airport (6 km to the southwest of ORS). The development dataset was sorted according to the nearest compass direction of the gradient wind. The median surface onshore component from ORS at 1400 EST was calculated, and then plotted, for each of the eight compass directions. Although the ORS onshore surface component does not vary greatly for onshore gradient winds, from the northeast to the southeast, there is more variation evident for offshore gradient winds. The median onshore surface components vary from −2.6 kn for a westerly gradient flow to 6.4 kn for an easterly flow. In order to gauge the diurnal contribution of the thermal forcing (sea-breeze and land-breeze/drainage flow) under each gradient wind direction, the onshore surface
component difference between 1400 EST and 0900 EST was calculated and plotted (Fig. 5, lower panel). In the median case, all directions registered a positive contribution to the onshore component between 0900 and 1400 EST. The largest change in surface onshore components was 10.6 kn for southwest gradient flows. Figure 5 (lower panel) clearly indicates that there is a large variation in surface wind behaviour, dependent on the gradient wind direction.

Most of the statistical techniques discussed below, with the exception of Stratified MOS for Sydney Harbour (SMOSSH), were developed from a learning sample of observed data. For operational use, each wind prediction scheme relied on gradient wind forecasts from NWP models. Further input such as estimates of cloud cover, temperatures and stability parameters was required for some of the techniques.

A gradient wind partitioned climatology-\textit{HarbClim}

A Sydney Harbour climatology (HarbClim) was developed using the surface winds from Fort Denison, the Offshore Reference Station, and Wedding Cake (see Fig. 1) – together with upper winds from Sydney Airport. The development database comprised of August, September and October observations for the years 1991-99. The data were partitioned according to the speed and direction of the 2300 UTC Sydney Airport 900 hPa wind. It would have been desirable to define a climatology for the eight compass points (Fig. 5), however, as the data (n=828) were also to be partitioned at 5 kn intervals, it would have been difficult to maintain partitioned datasets of a sufficient size. Following the conceptual model of Houghton (1992), a compromise was to stratify the data directionally into coastline-oriented quadrants according to the Sydney basin coastline orientation of around 20°. That is:

- if the gradient wind bearing lies between 300 and 020 then the quadrant allocated to the observation was NW.
- if the gradient wind bearing lies between 030 and 110 then the quadrant allocated to the observation was NE.
- if the gradient wind bearing lies between 120 and 200 then the quadrant allocated to the observation was SE.
- if the gradient wind bearing lies between 210 and 290 then the quadrant allocated to the observation was SW.

It is illustrative to look at the way that the surface wind at the Offshore Reference Station varies during the day by examining the onshore component of the wind (Fig. 6). In the median case, when all data is considered (Fig. 6(a)), there is an offshore flow until 1100 EST then the onshore component gradually builds until mid-afternoon. The range, or variation from the median is large. Fig. 6(b) displays only the partitioned data from the SW quadrant for gradient winds with wind speeds between 7.5 and 12.5 kn. The median case displays a similar progression of hourly winds to that shown in Fig. 6(a). However, the variation from the median is much less. In Fig. 6(c), where
only data from the SW quadrant with 900 hPa wind speeds between 17.5 and 22.5 kn are considered, the surface wind stays offshore all day in the median case. Data stratification on both direction and speed of the 900 hPa wind provide a more homogeneous dataset from which to derive a climatology.

The mean wind speed and modal direction for each hour were calculated from this historical database for each wind bearing and speed partition. For example, in HarbClim the climatology for NW 20 (for each station) was derived from all the observations with gradient wind speeds between 17.5 and 22.5 kn from the NW quadrant in the database. There is a variation about these climatological values dependent on the actual gradient wind, atmospheric stability, and any synoptic wind changes throughout the day. Forecasters were encouraged to make allowance for these factors.

Table 1 shows an example of HarbClim output for gradient winds of 20 kn from the NW quadrant. Useful information that can be gleaned from this partitioned climatology includes the most typical onset time of the sea-breeze, and the relative strengths of the winds in each location. Table 1 illustrates the typical behaviour of the sea-breeze in this gradient wind regime, becoming initially evident offshore (ORS) in the early morning and working its way inshore to Fort Denison in the early afternoon. The situation exemplified by HarbClim before 1000 EST, where a westerly surface flow is present in the inner harbour (Fort Denison) while a sea-breeze is indicated in the offshore areas and the middle harbour (Wedding Cake), was often observed during the morning hours as the sea-breeze front struggles to penetrate inland against an offshore flow.

### Stratified MOS for Sydney Harbour - SMOSSH

Similar to HarbClim, SMOSSH was developed on a dataset partitioned, or stratified, on the basis of the 900 hPa wind. However, rather than observed data, numerical model prognosis data were employed as predictors following the MOS methodology of Glahm and Lowry (1972). The development dataset consisted of predominantly LAPS (Puri et al. 1998) prognosis grid-point output from the 1200 UTC model run for the months August to November, 1993 – 1999. To ensure the stability of the wind prediction equations, it is desirable to have the number of daily wind data

---

**Table 1.** An example of HarbClim output for a gradient wind of 20 kn from the northwest. Hourly wind directions (degrees true) and wind speeds (kn) are given for two harbour locations (Fort Denison and Wedding Cake) and offshore (ORS).

<table>
<thead>
<tr>
<th>Time (EST)</th>
<th>Fort Denison</th>
<th>Wedding Cake</th>
<th>ORS-Offshore</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dir   Speed</td>
<td>Dir   Speed</td>
<td>Dir  Speed</td>
</tr>
<tr>
<td>0700</td>
<td>300   6</td>
<td>10    5</td>
<td>10   7</td>
</tr>
<tr>
<td>0800</td>
<td>300   5</td>
<td>20    6</td>
<td>30   8</td>
</tr>
<tr>
<td>0900</td>
<td>300   6</td>
<td>20    9</td>
<td>30   8</td>
</tr>
<tr>
<td>1000</td>
<td>310   9</td>
<td>350   11</td>
<td>40   10</td>
</tr>
<tr>
<td>1100</td>
<td>320   9</td>
<td>40    13</td>
<td>40   12</td>
</tr>
<tr>
<td>1200</td>
<td>80    11</td>
<td>40    14</td>
<td>50   12</td>
</tr>
<tr>
<td>1300</td>
<td>80    12</td>
<td>40    12</td>
<td>40   14</td>
</tr>
<tr>
<td>1400</td>
<td>80    12</td>
<td>40    13</td>
<td>50   14</td>
</tr>
<tr>
<td>1500</td>
<td>80    13</td>
<td>40    14</td>
<td>40   15</td>
</tr>
<tr>
<td>1600</td>
<td>80    13</td>
<td>40    16</td>
<td>50   15</td>
</tr>
<tr>
<td>1700</td>
<td>60    12</td>
<td>40    15</td>
<td>50   14</td>
</tr>
<tr>
<td>1800</td>
<td>50    11</td>
<td>40    14</td>
<td>40   16</td>
</tr>
</tbody>
</table>
values for each partition to exceed 200 (Wilson 1985). This condition necessitated the combination of the NE and SE quadrant data into a single 'Onshore' partition. Prediction equations were then derived on an hourly basis for the wind speed, onshore surface component, and parallel to shore component for each of the reference sites: Fort Denison, Wedding Cake and ORS. Although a large portion of the variance in the surface wind components can be explained by model component winds, temperature based potential predictors were also considered important. Previous studies (e.g., Lemcke and Kruizinga 1988; Roebber and Bosart 1996) demonstrate some skill gain in short-term temperature forecasting for experienced forecasters over objective NWP guidance. In an effort to capture this skill, some manually entered data was required by the operational forecasters prior to the running of SMOSSH – these were: sea-surface temperature (SST); maximum inland temperature (Richmond, 60 km NW of Sydney); and the inland temperatures at 0900, 1200, 1500 and 1800 EST. Other factors are also influential. A complete list of the potential predictors offered to the forward stepwise regression procedure (see Connor 1999 for further details) is shown in Table 2. Seasonal parameters were included due to the significant factor of sunset time on wind speed (Acevedo and Fitzjarrald 2001). The wind vector output from SMOSSH was converted to an hourly wind direction for each of the reference sites. The SMOSSH output from September 27 2000, using LAPS data from the 1200 UTC model run on the previous night, is represented in Table 3.

The manual input, derivation of special variables, and data restrictions caused by the stratification procedure are an increase in complexity when compared to the development and application of traditional MOS schemes that work objectively from numerical model data. In an effort to determine if

Table 2. The 76 potential predictors used in the stepwise linear regression procedures. The units for each category are shown in parentheses. The number of parameters in each category is also shown.

<table>
<thead>
<tr>
<th>Parameter categories</th>
<th>Description of potential predictors</th>
</tr>
</thead>
</table>
| Standard MOS         | Height (m) - for 1000, 950, 900, 850, 700 hPa  
                       | Temperature (K) - for 1000, 950, 900, 850, 700 hPa  
                       | East-west wind component (U) (m s⁻¹) - for 1000, 950, 900, 850, 700 hPa  
                       | North-south wind component (V) (m s⁻¹) - for 1000, 950, 900, 850, 700 hPa  
                       | Wind magnitude (m s⁻¹) - for 1000, 950, 900, 850, 700 hPa  
                       | Vertical velocities (W) (hPa h⁻¹) - for 1000, 950, 900, 850, 700 hPa  
                       | RH (%) - for 1000, 950, 900, 850, 700, 500 hPa  
                       | Vertical thickness (m) - for layer 1000/500 hPa  
                       | Precipitable water (m).  
                       | LAPS model precipitation (mm).  |
| Solar and Seasonal Predictors (3) | Sine and cosine of Julian day; Julian day.  |
| Sea-breeze and wind-derived Predictors (26) | Lapse rates (°C km⁻¹) - for layers 1000/900 hPa, 1000/850 hPa  
                       | Return Flow index - for layers 1000/850 hPa, 850/700 hPa  
                       | Product of the Lapse rate 1000/900 hPa with:  
                       | East-west wind component (U) (m s⁻¹) - for 1000, 950, 900, 850 hPa  
                       | North-south wind component (V) (m s⁻¹) - for 1000, 950, 900, 850 hPa  
                       | Wind magnitude (m s⁻¹) - for 1000, 950, 900, 850 hPa  
                       | [Wind magnitude]² (m² s⁻²) - for 1000, 900 hPa  
                       | [East-west wind component (U)]² (m² s⁻²) - for 1000, 900 hPa  
                       | [North-south wind component (V)]² (m² s⁻²) - for 1000, 900 hPa  
                       | Wind Bearing 900 hPa (Degrees) (derived from U 900 and V 900 hPa)  
                       | Deviation from median bearing in NW data set (Degrees)  
                       | Deviation from median bearing in SW data set (Degrees)  
                       | Deviation from median bearing in Onshore data set (Degrees)  
| Predictors derived from manually entered temperature data (8) | SST, TMAX - Maximum inland (Richmond) temperature (°C)  
                       | Maximum inland (Richmond) virtual temperature (°C)  
                       | Thermal Contrasts, virtual temperature difference between TMAX, T_09, T_12, T_15, T_18 and the SST.  
                       | Cube of Thermal Contrast (TMAX-SST)³ |
Table 3. The output from SMOSSH for 27 September 2000, using LAPS data from the 1200 UTC model run on the previous night. Hourly wind guidance for direction (WDIR) and speed (WSPD) is given for Fort Denison, Wedding Cake and ORS. The LAPS +12h forecast 900 hPa wind (in this case, bearing 344° at a speed of 10 kn) is displayed to give the forecasters an indicator of the model data from which the surface winds are derived. Output time is shown in EST, and the wind direction and speed are shown in degrees true and knots, respectively.

<table>
<thead>
<tr>
<th>Time</th>
<th>Fort Denison</th>
<th>Wedding Cake</th>
<th>Offshore (ORS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WDIR</td>
<td>WSPD</td>
<td>WDIR</td>
</tr>
<tr>
<td>0700</td>
<td>310</td>
<td>8</td>
<td>60</td>
</tr>
<tr>
<td>0800</td>
<td>310</td>
<td>9</td>
<td>50</td>
</tr>
<tr>
<td>0900</td>
<td>320</td>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>1000</td>
<td>350</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>1100</td>
<td>40</td>
<td>7</td>
<td>20</td>
</tr>
<tr>
<td>1200</td>
<td>40</td>
<td>9</td>
<td>40</td>
</tr>
<tr>
<td>1300</td>
<td>50</td>
<td>8</td>
<td>30</td>
</tr>
<tr>
<td>1400</td>
<td>80</td>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td>1500</td>
<td>60</td>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td>1600</td>
<td>90</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>1700</td>
<td>20</td>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>1800</td>
<td>60</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

the additional resources required to develop a stratified MOS scheme were worthwhile, the output from SMOSSH for the ORS site (corrected to 10 m due to a difference in mast height), and the MOS output for Sydney Airport (6 km to the southwest of ORS) were compared. The two statistical schemes were verified against Sydney Airport surface wind observations at the times 0900 and 1500 EST during a period that spanned the Olympic and Paralympic Games periods (3-30 September and 13-27 October 2000). The root mean square vector errors were calculated and then expressed as a Brier skill score with respect to climatology (BSSc) (e.g., Connor and Woodcock 2000).

Superior skill is displayed for the more protracted SMOSSH methodology over MOS methods for this data. When vector errors were considered, the BSSc for each method was 0.60 and 0.46, respectively.

Similar to Roeber (1998), in any regime-based statistical prediction scheme, the skill gains are thought to emerge by minimising the errors for a particular synoptic regime, rather than the traditional MOS approach where mean square errors are minimised for the developmental data as a whole. However, the additional skill does come with a greater development complexity that would seem only worthwhile for specific forecasting applications where small accuracy gains are justified.

A classification and regression tree (CART) technique - Tree Breeze

The strong dependence of the harbour surface winds on the gradient wind suggested the use of decision trees. Using the CART methodology (Breiman et al. 1984), a simple truncated tree with binary decision thresholds is developed by a series of univariate regression analyses. From a learning sample, the most influential predictor is screened and a decision threshold value is determined through cluster analysis (Hartigan and Wong 1979). This procedure is then repeated in a hierarchical fashion to create further nodes in the decision tree.

An example of a simple decision tree is shown in Fig. 7. This tree represents decisions needed to make Fort Denison 3 pm wind direction predictions based on a matched learning sample of 1991 to 1997 Fort Denison and Offshore Reference Station surface winds (from August to October) and Sydney Airport 900 hPa radiosonde data. The tree ‘yes’ branches split to the left, the ‘no’ branches split to the right. The predictions are found at the end points of the branches. Tree Breeze uses similar, but more complex, decision trees to give a forecast of the surface wind at hourly intervals throughout the day and uses additional parameters such as land/sea-surface temperature contrast, cloudiness, and upper wind speeds and directions (see Table 4).
Table 4. The 20 potential predictors offered to the CART algorithm that was used by Tree Breeze for predictor selection. The units for each category are shown in parentheses. The U and V components are defined relative to the coastline near Sydney Harbour.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPD9003AM</td>
<td>Wind speed at 900 hPa at 3am (m s(^{-1}))</td>
</tr>
<tr>
<td>SPD9009AM</td>
<td>Wind speed at 900 hPa 9am (m s(^{-1}))</td>
</tr>
<tr>
<td>SPD9003PM</td>
<td>Wind speed at 900 hPa 3pm (m s(^{-1}))</td>
</tr>
<tr>
<td>DIR9003AM</td>
<td>Wind direction at 900 hPa 3am (degrees)</td>
</tr>
<tr>
<td>DIR9009AM</td>
<td>Wind direction at 900 hPa 9am (degrees)</td>
</tr>
<tr>
<td>DIR9003PM</td>
<td>Wind direction at 900 hPa 3pm (degrees)</td>
</tr>
<tr>
<td>TMAXRICH</td>
<td>Maximum temperature at Richmond (°C)</td>
</tr>
<tr>
<td>TMINRICH</td>
<td>Minimum temperature at Richmond (°C)</td>
</tr>
<tr>
<td>TMINOBOSH</td>
<td>Minimum temperature at Observatory Hill (°C)</td>
</tr>
<tr>
<td>SST</td>
<td>Sea surface temperature at the Offshore Reference Station (°C)</td>
</tr>
<tr>
<td>UP3AM900</td>
<td>U component of the 900hPa wind at 3am (m s(^{-1}))</td>
</tr>
<tr>
<td>UP9AM900</td>
<td>U component of the 900hPa wind at 9am (m s(^{-1}))</td>
</tr>
<tr>
<td>UP3PM900</td>
<td>U component of the 900hPa wind at 3pm (m s(^{-1}))</td>
</tr>
<tr>
<td>VF3AM900</td>
<td>V component of the 900hPa wind at 3am (m s(^{-1}))</td>
</tr>
<tr>
<td>VF9AM900</td>
<td>V component of the 900hPa wind at 9am (m s(^{-1}))</td>
</tr>
<tr>
<td>VF3PM900</td>
<td>V component of the 900hPa wind at 3pm (m s(^{-1}))</td>
</tr>
<tr>
<td>MOL2AP0</td>
<td>Lapse rate from 850 hPa to Richmond temperature at 9am (°C km(^{-1}))</td>
</tr>
<tr>
<td>MOL2AP1</td>
<td>Lapse rate from 850 hPa to Observatory Hill temperature at 9am (°C km(^{-1}))</td>
</tr>
<tr>
<td>MOL2AP2</td>
<td>Lapse rate from 850 hPa to Offshore temperature at 9am (°C km(^{-1}))</td>
</tr>
<tr>
<td>THERCON2</td>
<td>Difference between maximum temperature at Richmond and the SST (°C)</td>
</tr>
</tbody>
</table>

Fig. 7. An example of a simplified recursive partitioning tree, similar to the final tree used for determining the 3 pm surface wind direction at Fort Denison using the 900 hPa wind speed and direction as predictors. Decision nodes are located at each intersection. If the decision rule at a given node is true, cases follow the left branch; if false, cases follow the right branch. The number below each terminal node represents the average wind direction for cases assigned to that node. The abbreviations used at each decision threshold are expanded upon in Table 4.
The CART software used in this study (S-Plus) is specifically designed for this type of data analysis and will prune back a very large tree to an optimum size, and allow error analysis fairly simply. The software has the advantage that output can be easily translated into Java script and was used to develop the final trees used for the Olympic Games in 2000. Experimentation with CART showed that it was better to predict the U and V components of the surface winds, and then derive wind speed and direction from them, as this avoids decision problems that occur at the end points of 0 and 360 degrees. The method will select the optimum predictors out of a large potential predictor set. Table 4 shows the predictors offered to CART to develop the final trees.

For example, to develop a relationship for the 3 pm U component of the surface wind direction at Fort Denison, the CART software developed a set of ‘IF’ statements based on only six of the 20 potential predictors shown in Table 4. The optimal regression tree proposed by CART has 16 nodes (‘IF’ statements), and was pruned back from an extremely large potential tree. This is shown conceptually in Fig. 8, where a minimum relative error of 0.478 occurred at 16 nodes.

CART can be used to derive yes/no results for the occurrence of a sea-breeze, as well as to derive estimates for the wind speed and direction. Initial investigations using observed gradient winds and temperature data, gave results on independent 1998 test data of a probability of detection (POD) of 82% and a probability of false detection (POFD) of 26%. This is comparable with the results of the logistic regression methodology described in detail in Dunsmuir et al. (2003), and briefly, below. The logistic regression gave a POD of 80%, and a POFD of 22% on the same data.

By applying this methodology to the historical dataset for the U and V components of the surface wind at Fort Denison, for each half hour between 9 am and 6 pm, and smoothing the resulting data, a useful prediction system was obtained for wind speed and direction. Similar equations were obtained for the Offshore Reference Station. At other harbour locations, the development data set was considered too short to apply this method.

A combined logistic and linear regression technique - Regression Breeze
A yes/no forecast for the sea-breeze occurrence at Fort Denison was developed using logistic regression on a number of variables relevant to the sea-breeze, similar to those described in the previous sections and based on the work of Connor (1997). This is described in Dunsmuir et al. (2003). This was a first step in developing a multiple linear regression model to predict sea-breezes. As discussed in this paper, scatter plots of the surface wind at Fort Denison versus the gradient wind showed three different regimes that needed to be treated separately in the linear models:
1. the gradient wind offshore, but the surface wind a sea-breeze;
2. the gradient wind and the surface wind both onshore; and
3. the remainder.

The process is shown in the flow chart in Fig. 9.

Logistic regression was used to subdivide days with an offshore gradient wind into days of sea-breeze occurrence and non-occurrence, thereby, dividing the offshore gradient wind days into categories 1 and 3 above. During the period, for which the analysis was performed, it was found that about 80% of days had an offshore gradient wind, with 50% of these having a sea-breeze occurring. Multiple linear regression was applied to each of the categories to derive equations for the Fort Denison and Offshore Reference Station surface wind speed and direction.

The regression equations were initially developed on data from August, September and October of 1991 to 1997 (Kim 1999). For Fort Denison, the linear regression schemes for the onset time gave a root mean square error (RMSE) of 1 hour on the developmental data set and a RMSE of 1 h 30 min on the independent 1998 data when observed upper winds were used as predictors. For the sea-breeze at 3 pm (linear regression case 1), wind speed and direction, the RMSE were 3.7 kn and 34 degrees; for the ‘no-sea-breeze’ case at 3 pm (linear regression case 3), these RMSE values were 3.9 kn and 111 degrees; for the onshore gradient wind situations at 3 pm (linear regression case 2), the RMSE values were 3.5 kn and 35 degrees.

The regression-based method was tested operationally during the 1999 Olympic Test Events, held
from 16 to 27 September 1999. Overall, for this test event, RMSE values were found to be 40 degrees for direction and 2.9 kn for speed, calculated between the hours of noon and 5 pm. These results were obtained retrospectively, and based on observed upper winds rather than forecast winds. This was necessary because the method proved to be operationally impractical as too many variables had to be manually entered, and forecasters did not have the time available to apply the method daily. However, these results, on a limited data period, encouraged the use of a simplified version of this method, called Regression Breeze, with reduced parameter inputs, as part of the Olympic prediction suite. New equations were prepared which included an additional year’s data, and which differed from the original equations primarily in two ways:

1. The number of variables was restricted to ensure ease of data input by operational forecasters, as described in Dunsmaur et al. (2003).
2. Predictions were made for the U and V components of the wind. These were then combined to create the wind speed and direction.

A wind gust prediction technique - Gustfactor

The statistical techniques previously discussed allow the derivation of relationships between the 10-minute average wind and a range of meteorological parameters. Wind gusts, or peak wind speeds of a short duration, are also important for the planning of a yachting regatta. Sydney Harbour lies in a region of complex topography (Fig. 1), and the wind flow is usually turbulent and gusty – particularly from directions that include intricate terrain or urban landscape.

Previous studies (e.g., Hart and Forbes 1999) have explored the transport of high-momentum air to the surface and noted the sensitivity of wind gusts to boundary layer stability and vertical wind shear. Fujita and Wakimoto (1982) and Acevedo and Fitzjarrald (2001) demonstrate that sheltering and turbulence effects are site specific and dependent on wind direction. Acevedo and Fitzjarrald (2001) also noted temporal influences on gustiness, associated with surface sensible heat flux.

Although lower tropospheric stability is obviously important in turbulent eddy transfer of upper winds to the surface, stability is difficult to operationally forecast on an hourly basis. Surface temperature was thus used as a proxy for stability in the developmental database for Gustfactor. Historical wind gust data sets (1993-99, July to November) for Fort Denison and Wedding Cake were examined for the influence of air temperature, time of day and wind direction on gustiness. Parameters based upon the wind direction explained the most variance when relationships between gusts and wind speed were derived, and the wind direction was subsequently used as a basis for partitioning the data at 10 degree intervals. Non-linear regression was used to derive prediction equations for wind gusts. Direct relationships between wind magnitude and gusts were also determined. The best results were achieved with a bivariate relationship using the mean wind speed together with the square root of the wind speed. Gustfactor was presented in tabular form to provide an easy to use operational guide to wind gusts – once the forecast mean wind and direction, for a given location, had been determined. Forecasters were encouraged to judge the Gustfactor output slightly according to their determination of low-level stability. Fig. 10 compares the observed and predicted values for Gustfactor output for Fort Denison when the wind direction is 240°.

The gust prediction equation derived from the data displayed in Fig. 10 explains 98% of the variance of the wind gusts and is typical of all the direction-based equations. As an example, the relationship between

---

**Fig. 10** Observed and Gustfactor predicted wind gusts for Fort Denison when the wind bearing is 240 degrees. The Pearson R correlation was 0.99 for n=91.
wind gusts and wind speed from a bearing of 240 degrees at Fort Denison can be represented

\[
\text{Wind Gust}_{240} = 1.52 \times \text{Wind Speed} + 1.18 \times \text{Wind Speed}^{0.5}
\]

Upstream obstruction of the wind flow was influential and the wind directions that were most gusty were 240° and 040° (see Fig. 1) with a ratio of maximum gust to mean wind speed of 1.89 and 1.69, respectively. The least gusty directions of 100° and 290° were aligned with the harbour shoreline, with gust ratios of 1.34 and 1.47, respectively.

**Summary and conclusions**

The statistical tools described in this paper, together with a high resolution NWP model over the Sydney domain, were designed to provide guidance to the 2000 Olympic and Paralympic Games. Meteorologists would base their forecasts on this guidance and adjust the output for any changes not defined by the model, such as an unforeseen synoptic change, or any systematic bias displayed by the guidance. Early versions of some of the statistical tools were tested during the Olympic trial event held in Sydney Harbour during September 1999. Spark and Connor (2003) demonstrate that the suite of statistically-derived products described here were a vital part of the Bureau of Meteorology contribution to wind forecasting during the Olympic and Paralympic games, where all the statistical models were utilised as real-time prognosis products.

The statistical models were designed to be used in conjunction with high resolution numerical models over the Sydney domain to provide a full spectrum of guidance products for the Olympic operational meteorologists. The advantage of statistical products for fine-scale forecasting is their ability to link observations with numerical model output. Local effects caused by detailed topography are automatically accounted for by this process. Spark and Connor (2003) provide a verification of the statistical tools described in this paper and a comparison, for wind forecasting skill, against numerical model output and the official Sydney Olympic Sailing Weather Office (SOSWO) forecasts.

The statistical forecasting methodologies described in this paper, such as the selection of meteorological parameters pertinent to the low-level flow (in particular, the sea-breeze), and the partitioning of development data according to synoptic characteristics, display promise in obtaining skill gains for coastal wind forecasting. In the Olympic trial events, the Regression Breeze model showed solid predictive potential in a time-lagged trial using observed gradient winds. Increased skill is also demonstrated by the SMOSSH methodology over the more conventional MOS methods under operational conditions. However, in each case, the additional skill is associated with an increased development complexity. The more intricate methodology would seem most appropriate for special forecasting applications, or, for forecasting parameters routinely where small skill increases would prove beneficial.

**Acknowledgments**

The work of Robert Dahni is acknowledged in the development of SMOSSH. His IDL programming skills were integral to the implementation of this prediction scheme. Many thanks are also due to Sung Kwan Kim and Shuiling Chen for their help in developing the data set used in the development of these statistical wind prediction methods for Sydney Harbour. The initial stimulation to explore the techniques described in this paper was provided by the earlier, unpublished, work of Ken Batt and Bruce Buckley, from the Bureau of Meteorology, on the winds of Sydney Harbour. Thanks to Jim Fraser for his model data extraction skills. Thank you to Samantha Taylor from the Severe Weather section of the Bureau of Meteorology, Queensland, and the two referees for their detailed comments on earlier drafts that have helped to improve the clarity of this paper.

**References**


