# Overall Methodology

## Monitoring data used

All available criteria pollutant data, including IMPROVE visibility monitors at wilderness areas and some sites that are not submitted to AQS, were used to maximize the spatial distribution of monitors. A few sites that commenced partway through the 2009- 2011 period, or were discontinued after at least one year of data within this 3-year time window, were used. The website shows all monitors used in this exercise, on maps of the respective pollutant concentrations.

The model - monitor interpolation process requires the presence of a few monitors just outside the three states, to ensure that interpolated values closer to state borders remain realistic. Data from monitors in southern British Columbia, northern California, Nevada and Utah, and western Wyoming and Montana were used as available. A few fictitious marine monitors (assigned the same background design values as recorded at rural coastal sites) were also placed in the Pacific Ocean.

The algorithm for interpolating modeled and monitored data is available for easy use in a separate EPA model (BenMap- US EPA, 2008). It is important that no monitors outside the modeling domain are used, as this leads to large errors in BenMap's interpolation.

Monitor data treatment specifics:

- When multiple non-FRM monitors are available at the same site, the following order of preference was followed and data substituted accordingly: FEM > Neph > TEOM. If an FRM monitor was operated, its design value was used directly, with no collocated data substitution on the days it did not run.
- Corrected baseline drift in CO between Sept- Nov 2011, at Custer, WA and removed a very high single-hour spike
- Removed baseline drift in $SO_2$ between Jan 2009- 16 Sept 2010, and after that until Dec 2011, at Anacortes, WA
- Cheeka Peak $SO_2$: hour during which calibration checks were run were not appropriately flagged in 2009- 2010. These hours were removed from consideration.

- Some Canadian sites (those operated by MetroVancouver) report data at the ending hour instead of the starting hour, which is the EPA convention used for all other sites. Data from these sites were relabeled one hour backward.
- Several BC Ministry of Environment sites reported $PM_{2.5}$ and $PM_{10}$ concentrations under standard conditions (25ºC and 1atm) through early 2011. After obtaining the changeover dates, meteorological data from nearby airports were used to convert to ambient conditions.
- Omitted data from $SO_2$ monitors at Trail, BC and Pocatello, ID. Though these are ambient monitors, they are located very close to significant $SO_2$ sources.

Design values for each criteria pollutant (other than Pb) were calculated following the rules at http://www.epa.gov/air/criteria.html. However quarterly data completeness requirements were sometimes relaxed to maximize the monitoring data available, as long as it did not over-represent some seasons or unduly bias the final product. No data from exceptional events, whether flagged or not, were excluded in this exercise. Data values, either including or excluding "exceptional event" days flagged by state agencies may be obtained at EPA's AirData website (http://www.epa.gov/airdata/). "Exceptional events" have not necessarily been sanctioned by EPA, so for critical modeling projects such as Prevention of Significant Deterioration (PSD) permitting, please consult with your state monitoring and modeling personnel to determine officially concurred values for any analysis that deviates from the values produced by this tool.

**Modeling data used**

Washington State University runs a gridded photochemical atmospheric dispersion model, the Community Multi-Scale Air Quality Model (CMAQ), in predictive mode with 12km grid spacing, every day (http://airpact-3.wsu.edu). Predicted meteorological fields on the same grid are available twice daily from the University of Washington's WRF model (http://www.atmos.washington.edu/mm5rt). A separate gridded emissions module incorporates statewide land use patterns, traffic volumes, industrial and biogenic emissions, all adjusted as appropriate by season, day of week, time of day and predicted temperature and solar radiation. Washington Department of Ecology maintains several years of archived model data covering Washington, Oregon and Idaho. Thus with the use of model data, this process is able to account for the effects of terrain, meteorology, emissions and photochemistry on the spatial variability of pollutant levels.

Model data treatment specifics:

- During archival of daily forecasts, hourly data is aggregated over a day to reduce data storage requirements. $PM_{2.5}$, $PM_{10}$, CO, $SO_2$ and $NO_2$ daily means, maximum daily 8-hour $O_3$ averages and daily maximum 1-hour $NO_2$ data were archived.
- Unfortunately neither of the hourly maximum CO, 8-hourly maximum CO, hourly maximum $SO_2$ or 3-hour maximum $SO_2$ concentrations were archived until 2012. Since both CO design values associated with hourly and 8-hourly standards are based on the second highest annual values, the modeled second-highest daily CO mean concentration was calculated for each year. The $SO_2$ design values associated with the hourly and 3-hourly standards are based on the annual 99$^{th}$ percentile and 2$^{nd}$ highest values respectively. The modeled second-highest daily $SO_2$ mean concentration was used in their place. For continuous, year-round monitoring, the second highest value would be higher than the 99$^{th}$ percentile (4$^{th}$ highest).
- The model's inclusion of wildfires is sporadic, and there were periods where it constantly over-predicted $PM_{2.5}$ during wildfires. Therefore for interpolating 24-hr $PM_{2.5}$ design values, modeled 3-year medians rather than a 3-year average of annual 98$^{th}$ percentiles were used.
- Gridded design values for the remaining pollutants and their appropriate averaging periods were calculated from model data the same way the monitoring data were treated.

## Spatial interpolation of monitor data with model data

The spatial interpolation was performed using the monitor to fix the value, and the monitor/ model ratio to spatially scale the data. The Voronoi nearest neighbor method was used to identify which monitors could be considered as nearest neighbors. The interpolation process is further described with an example in Appendix 1.

Even though the AIRPACT3 model incorporates the effects of terrain, the spatial interpolation scheme does not respect such barriers. As such two monitors on either side of the Cascades could be considered nearest neighbors and influence the interpolation of the nearby grid cells. This is particularly problematic when dealing with pollutants such as $PM_{2.5}$, which is often elevated due to the accumulation of woodsmoke in small valley communities. To prevent the $PM_{2.5}$ monitors from influencing the interpolation over unreasonably large areas, and due to the sufficiently dense network of $PM_{2.5}$ monitors, the monitored design values were downweighted more heavily as the distance from the monitor increased. An inverse distance squared ($1/D^2$) weighting function was used for $PM_{2.5}$, while 1/D weighting was used for all other pollutants.

Since the dispersion model (AIRPACT3) is used to spatially scale the monitor data (i.e., is a grid cell higher or lower than its neighbor?), the process assumes that the dispersion model captures large scale concentration gradients reasonably well. As such the interpolation is somewhat tolerant of inaccurate predictions of pollutant concentrations at monitors, because it closely reproduces the concentration of the monitor in the corresponding grid cell. The monitor's influence on the interpolation extends a little beyond its immediate grid, depending on the distance weighting used.

The sensitivity of the interpolation to the presence of monitors was examined with two approaches:

1. <u>By randomly omitting 10% of the $PM_{2.5}$ monitors (the densest monitoring network) and comparing the interpolated solutions.</u>

   Figure 3 shows such a difference plot.

   As can be seen, the interpolation is more sensitive to the omission of some randomly chosen sites than others. For the most part, differences do not extend far beyond the omitted site, except in the case of Lancaster, ID and to a lesser extent, in Mount Vernon, WA. The reasons for these differences are found in the relationship between the model and nearby monitors

   Figure 4 shows the location of Lancaster monitor (top left) and the Pinehurst, ID monitor (right). As can be seen, the model slightly overestimates the concentration at Lancaster, while largely underestimating that of Pinehurst. The latter is a small confined valley community of $<4km^2$ which is not resolved by the 12km AIRPACT3 model. Emissions are smeared over a larger area, leading to lower predicted concentrations. The model/ monitor ratio in this grid cell is 11/40. Lancaster is a larger community and the model/ monitor ratio is 21/12. The reason for this model over-prediction is thought to be an inconsistency in some of the emissions being modeled. In addition a monitor 36km to the west, in Spokane, WA was also omitted during this analysis, but as the model performance is within 4% of the monitored design value there, this is not expected to be the cause of the differences seen in Figure 3.

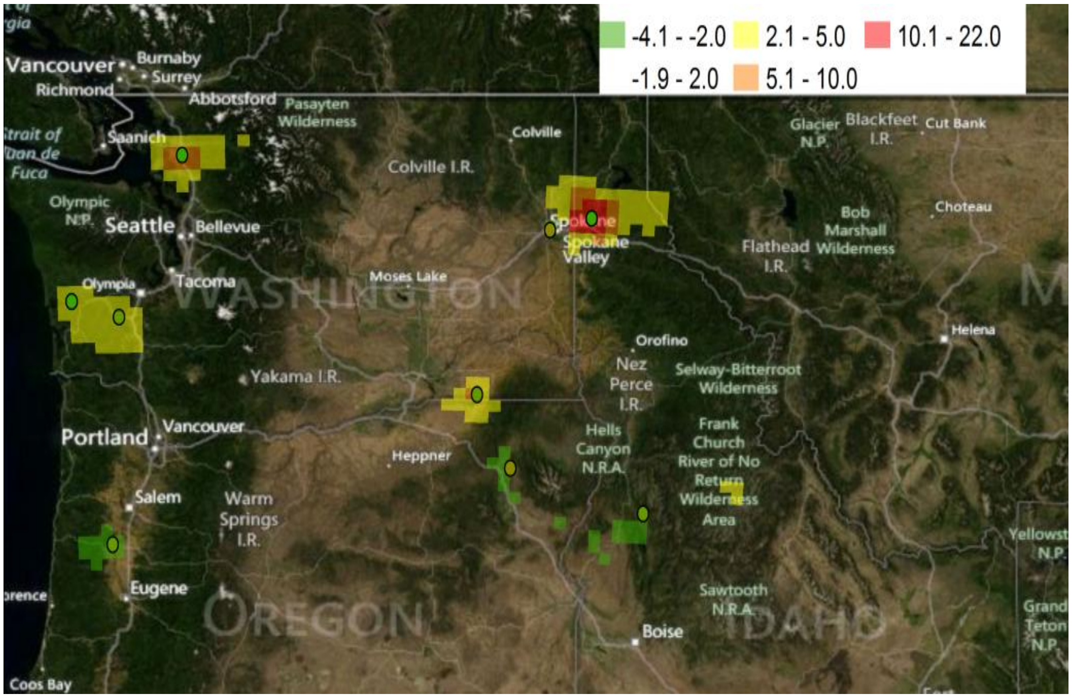



*Figure 3: Map of change in $PM_{2.5}$ 24-hr design value by omitting 10% randomly chosen sites. Monitors removed are shown as circles color coded by their design values. Grids within ±2 μg/m³ are not shown.*

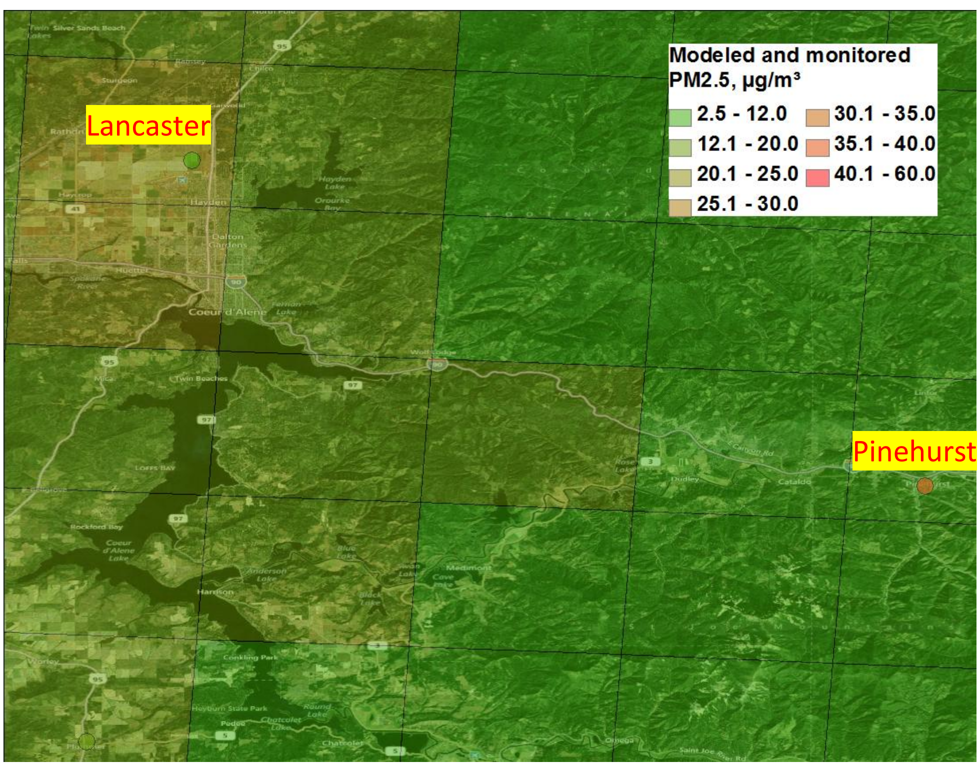



*Figure 4: Lancaster and Pinehurst modeled (areal shading) and monitored (filled circles) design values.*

The interpolation increases the concentrations in Pinehurst's grid by almost a factor of 4 to make it agree with the monitor, although the monitor's weight wanes as a function of inverse distance squared. About 4 grid cells (~50km) to the west in Lancaster, the interpolation has to downscale the modeled concentration by almost a factor of two. In terms of an analogy, this is akin to planting a tent peg at the west end while inserting a pole at the east end of the tent.

If the monitor at Lancaster is omitted from the interpolation, the tent lacks an anchor on the west side and the tarpaulin is left "fluttering in the breeze", driven mostly by the model's overestimate. This results in the intermediate areas showing high concentrations. This consequence is a limitation of both the model resolution and the $1/D^2$ weighting, in that the latter cannot be applied more tightly to some areas.

At Mount Vernon, the AIRPACT3 model overpredicts $PM_{2.5}$ by about 70%. Omitting this site from the interpolation allows the "tarpaulin to flutter" and artificially increases the concentration nearby.

2. By running the interpolation of $PM_{2.5}$ and $PM_{10}$ with and without the IMPROVE monitors.

Figures 5 and 6 compare $PM_{2.5}$ and $PM_{10}$ 24-hour design values at IMPROVE sites respectively, against the same values interpolated without including the IMPROVE monitors. $PM_{2.5}$ design values appear to be slightly overestimated at most IMPROVE sites, while they appear to be both under and overestimated in the case of $PM_{10}$. Two sites with high $PM_{10}$ design values, Three Sisters, OR (THSI) and Wishram, WA (CORI) are both driven by three high-impact days in the 3-year period (see Figure 1). Overall, the omission of these IMPROVE sites do not appear to cause very large errors in the interpolated concentrations.

AIRPACT3 model performance for ozone was better than that of other pollutants, so the interpolation will not be influenced much by the omission of sites. The sensitivity of the interpolation to the omission of CO, $NO_2$ or $SO_2$ monitors was not investigated due to the few monitors available within the 3 states.

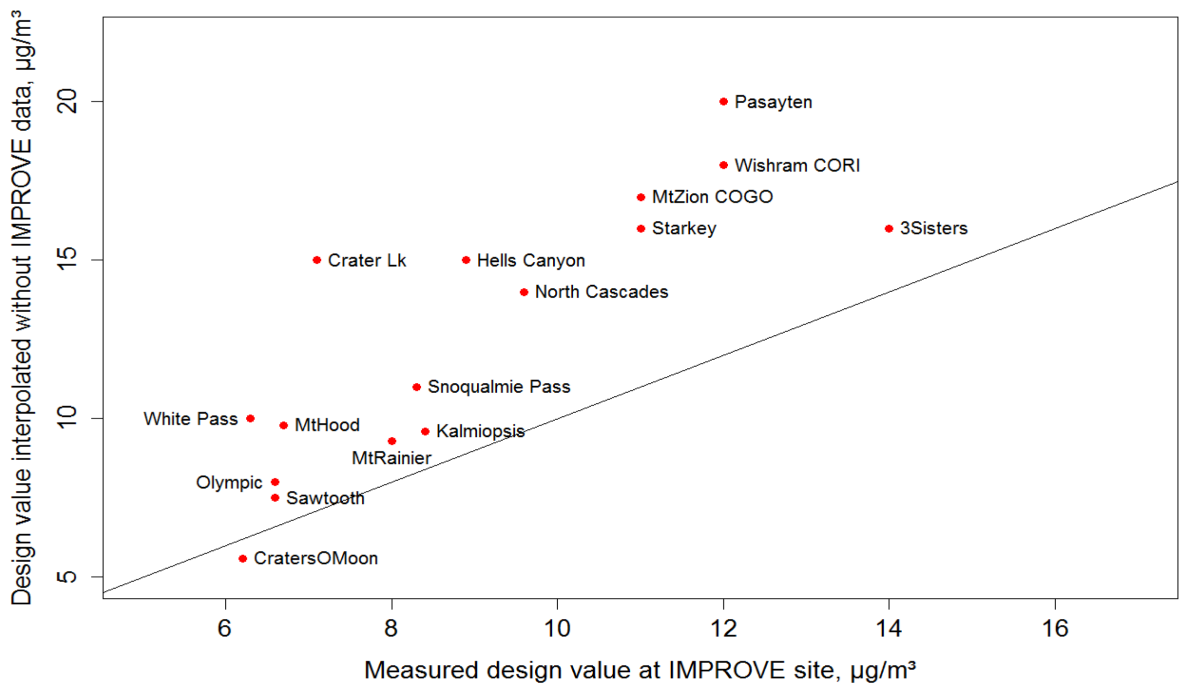



*Figure 5: $PM_{2.5}$ 24-hour design value interpolation sensitivity to the omission of IMPROVE data*

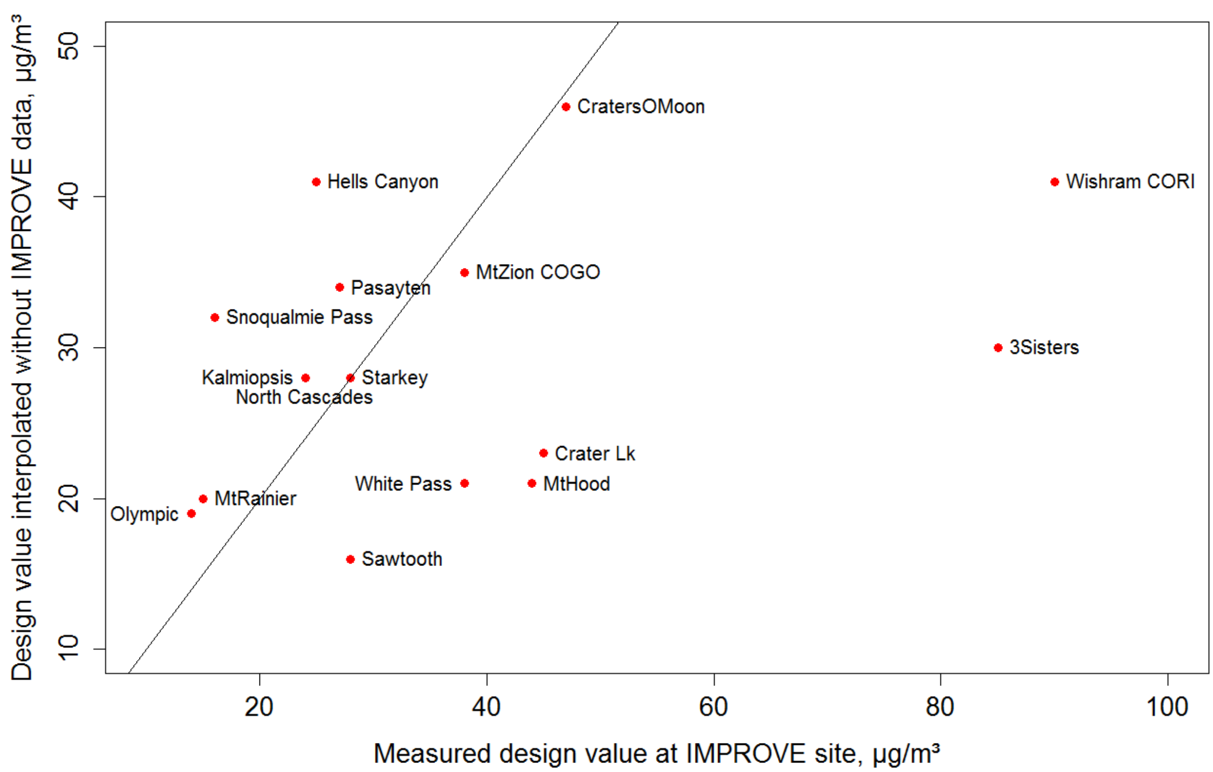
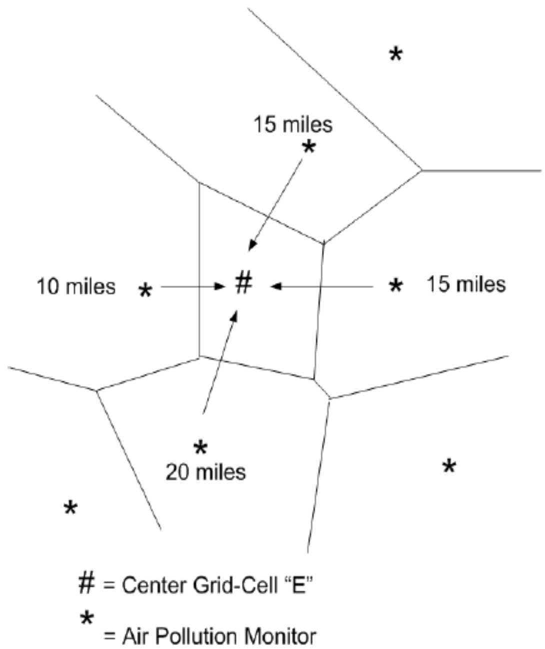



*Figure 6: $PM_{10}$ design value interpolation sensitivity to the omission of IMPROVE data*

**Appendix 1: Spatial interpolation of nearest neighbor monitors**

1. Drawing a polygon, or "Voronoi" cell, around the center of each grid cell.

2. Then choose those monitors that share a boundary with the center of grid-cell "E." These are the nearest neighbors. The relative weights of each monitor are based on their distances from the center of grid cell "E".




Relative weight of a monitor 10 miles away (using inverse distance squared or $1/D^2$ weighting) is given by:

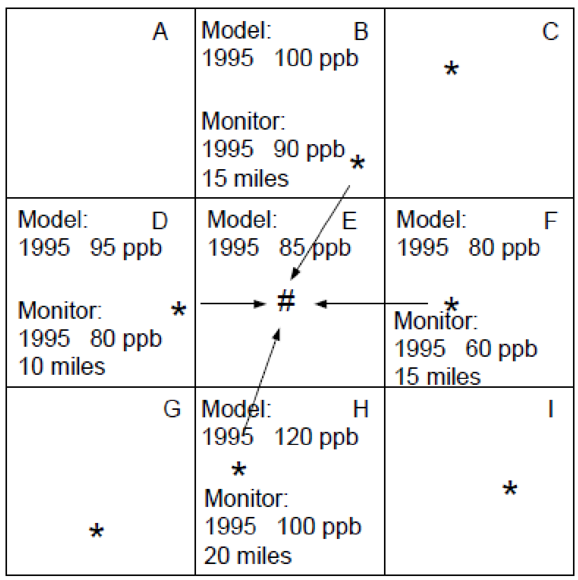
$$\frac{1/10^2}{\left(1/10^2\right)+\left(1/15^2\right)+\left(1/15^2\right)+\left(1/20^2\right)}$$

| Distance from grid center to monitor | 1/D relative weight | $1/D^2$ relative weight |
|---|---|---|
| 10 | 0.35 | 0.47 |
| 15 | 0.24 | 0.21 |
| 20 | 0.18 | 0.12 |

For pollutants likely to be concentrated into confined areas ($PM_{2.5}$), exponent= 2. For all other pollutants the exponent was set to 1. BenMap allows only a single exponent for the whole domain. There is no way to more forcibly constrain the pollutant concentrations recorded by some monitors (e.g., those in small valley communities) while allowing other monitors to influence a larger area.

3. Now use the monitor/model ratio for the nearest neighbors weighted by their respective distances, to scale the modeled value at grid cell "E". This spatial interpolation is repeated for each grid cell in the domain.

$$\left(0.35*80*\frac{85}{95}\right)+\left(0.24*90*\frac{85}{100}\right)+\left(0.24*60*\frac{85}{80}\right)+\left(0.18*100*\frac{85}{120}\right)=70.8\,ppb$$




# = Center Grid-Cell "E"

* = Air Pollution Monitor

0.1ppb more if interpolated with $1/D^2$ weighting

The value is interpolated to the center of the grid cell and represents the whole 12km x 12km area, such that there is no spatial variability within the $144km^2$ of the cell. This is a limitation of the model resolution. Even if a small community only covers a fraction of the overlapping grid cell, the concentration of that community's monitor will be assigned to the entire grid cell. In these areas, alternative analyses may yield more representative design values, but any alternative analysis should be discussed in advance with your state permitting contact.