Say there is a 2D continuous field $p(\vec{x},t)$ that fluctuates randomly in time. All we know about the variation is that the average is zero $\langle p(\vec{x})\rangle = 0$ and that the covariance between the field at two points is determined by the distance between them $\langle p(\vec{x})p(\vec{y})\rangle = f(|\vec{x}-\vec{y}|)$. the angled brackets $\langle\rangle$ denote the time average. As far as we are concerned each point in time is independent from all others.

There is a set of stations at positions $\{\vec{x}_i\}$ taking measurements $\{P_i(t)\}$ of the field. But the measurements are not entirely accurate - they are off by the random errors $\{\delta_i(t)\}$ such that at each time

$$P_i(t) = p(\vec{x}_i,t) + \delta_i(t) \tag{1}$$

Assume that the errors are independent $\langle\delta_i\delta_j\rangle = \langle p(\vec{x})\delta_i\rangle = 0$ and have zero mean $\langle\delta_i\rangle = 0$.

This implies

$$F_{ij} \equiv \langle P_i P_j\rangle = \langle p(\vec{x}_i)p(\vec{x}_j)\rangle = f(|\vec{x}_i - \vec{x}_j|) \tag{2}$$

> *My question is how to best predict the true field $p$ only using the measurements $P_i$.* Yes, I know the notion of "best" is yet to be defined.
>
> I first thought about this problem when thinking of ways to obtain accurate pressure gradients from real-world measurements of pressure to then obtain geostrophic wind which would be compared to the wind measured at those same stations. The common method is to triangulate the stations and find the gradient in each triangle. The problem was that small errors (which for pressure are quite comparable to the size of the variation) greatly exaggerated the estimates of the gradient and I thought I could do better.
>
> Here is what I came up with.

I propose to vary the prediction $q(\vec{x})$ to minimise the mean squared difference (wow, how original!) $\Delta(\vec{x}) = \langle(q(\vec{x})-p(\vec{x}))^2\rangle$ at all points. And I propose the prediction to be of the form

$$q(\vec{x},t) = \sum_i g_i(\vec{x})P_i(t) \tag{3}$$

where $\{g_i(\vec{x})\}$ are coefficients that do not vary with time. This makes the prediction at each point just a linear combination of the measurements at nearby stations weighted according to how much we trust each station's prediction. Trust can decrease both due to large separation and due to large errors.

With this setup $\Delta$ is straightforward to minimise. Choosing a point $\vec{x}$ and a time $t$ I find

$$\Delta - \langle p^2\rangle = \langle(q-p)^2\rangle - \langle p^2\rangle = \langle q^2\rangle - 2\langle qp\rangle = \left\langle \sum_{ij} g_i g_j P_i P_j \right\rangle - 2\left\langle \sum_i g_i p P_i \right\rangle = \sum_{ij} g_i g_j F_{ij} - 2\sum_i g_i f_i \tag{4}$$

where $f_i \equiv \langle pP_i\rangle = f(|\vec{x}-\vec{x}_i|)$. Now find the stationary point which is necessarily a minimum

$$\frac{\partial\Delta}{\partial g_i} = 2\sum_j g_j F_{ij} - 2f_i = 0 \tag{5}$$

In matrix notation

$$\mathbf{f}(\vec{x}) = F\mathbf{g}(\vec{x}) \tag{6}$$

We just have to invert the matrix to solve for the optimal coefficients

$$\mathbf{g}(\vec{x}) = F^{-1}\mathbf{f}(\vec{x}) \tag{7}$$

$$g_i(\vec{x}) = \sum_i (F^{-1})_{ij} f_i(\vec{x}) \tag{8}$$

The existence of the inverse may or may not be guaranteed.

Finally,

$$q(\vec{x}, t) = \mathbf{P}^T(t)\, F^{-1}\, \mathbf{f}(\vec{x}) \tag{9}$$

To make this approach work for real-world applications you would need to obtain $F$ and $f(r)$ from your data. The matrix $F$ is trivial, but $f(r)$ may cause problems. You could simply fit some function to the elements $F_{ij}$, but it would contradict equation 2. This shouldn't cause the method to break or anything, but still eww...

Have you seen something like this before? No way that I am the first to think of this. Or maybe there's a good reason people discarded it.

The neat thing is that if we instead minimise the squared differences in **gradient** we still obtain the same results.

Once again predict $q(\vec{x}, t) = \mathbf{P}^T(t)\mathbf{g}(\vec{x})$, and minimise $\Delta' = \langle(\nabla q - \nabla p)^2\rangle$.

So this is an example of non-parametric regression you say. The choice of kernel is not important you say. Only the bandwidth is you say. You can do cross-validation you say...

For the calculation of geostrophic wind $\vec{v} = \frac{\hat{\vec{R}}\times\nabla p}{2\rho\omega\cos\theta}$ you need to estimate the gradient of the pressure. Once again predict $q = \mathbf{P}^T\mathbf{g}$, and minimise $\Delta = \langle(\nabla q - \nabla p)^2\rangle$. $\nabla q = \mathbf{P}^T\nabla\mathbf{g}$.

$$\Delta - \langle(\nabla p)^2\rangle = \langle(\nabla q)^2\rangle - 2\langle\nabla p\cdot\nabla q\rangle = \langle\mathbf{P}^T\mathbf{P}^T\rangle\nabla\mathbf{g}\cdot\nabla\mathbf{g} - 2\langle\mathbf{P}^T\nabla p\rangle\cdot\nabla\mathbf{g}$$

Note that $\cdot$ is the inner product in the 2D real space, not the station space. Use $\langle\mathbf{P}^T\mathbf{P}^T\rangle = F$ and $\langle\mathbf{P}^T\nabla p\rangle = \nabla\mathbf{f}$ to get

$$\Delta - \langle(\nabla p)^2\rangle = F\nabla\mathbf{g}\cdot\nabla\mathbf{g} - 2\nabla\mathbf{f}\cdot\nabla\mathbf{g}$$

$$\frac{\partial\Delta}{\partial\nabla\mathbf{g}} = 2F\nabla\mathbf{g} - 2\nabla\mathbf{f} = 0$$

The result

$$\nabla q = \mathbf{P}^T F^{-1}\nabla\mathbf{f}$$

is the same as obtained before. So $\Delta = \langle(q-p)^2\rangle$ and $\Delta = \langle(\nabla q - \nabla p)^2\rangle$ are equivalent.

If we are interested only in the speed of the geostrophic wind we may want to choose $\Delta = \langle(|\nabla q| - |\nabla p|)^2\rangle$ or even have to involve the residual $p_0$ to account for dominating winds. In the treatment above however a nonzero mean wind has no influence. This would likely wouldn't be as nice. It might even need additional assumptions to calculate covariance.

An interesting side note is

$$\langle\nabla p(\vec{x})\cdot\nabla p(\vec{y})\rangle = -f''(|\vec{x}-\vec{y}|)$$

Proof:

$$\begin{aligned}
\langle\vec{n}\cdot\nabla p(\vec{x})\ \ \vec{n}\cdot\nabla p(\vec{y})\rangle &= \left\langle\lim_{\delta x,\delta y\to 0}[p(\vec{x}+\vec{n}\delta x) - p(\vec{x})][p(\vec{y}+\vec{n}\delta y) - p(\vec{y})]/\delta x\delta y\right\rangle \\
&= \lim_{\delta x,\delta y\to 0}[f(r+\vec{n}\cdot\hat{\vec{r}}(\delta x-\delta y)) - f(r-\vec{n}\cdot\hat{\vec{r}}\delta y) + f(r) - f(r+\vec{n}\cdot\hat{\vec{r}}\delta x)]/\delta x\delta y \\
&= \lim_{\delta y\to 0}\vec{n}\cdot\hat{\vec{r}}[f'(r-\vec{n}\cdot\hat{\vec{r}}\delta y) - f'(r)]/\delta y \\
&= -(\vec{n}\cdot\hat{\vec{r}})^2 f''(r)
\end{aligned}$$

, where $\vec{r} = \vec{x}-\vec{y}$, $r = |\vec{r}|$, and $\hat{\vec{r}} = \vec{r}/r$.

$$\begin{aligned}
\langle\nabla p(\vec{x})\cdot\nabla p(\vec{y})\rangle &= \sum_k\langle\vec{e}_k\cdot\nabla p(\vec{x})\ \ \vec{e}_k\cdot\nabla p(\vec{y})\rangle \\
&= -\sum_k(\vec{e}_k\cdot\hat{\vec{r}})^2 f''(r) \\
&= -f''(r)
\end{aligned}$$

□

$\langle\mathbf{P}^T\nabla p\rangle = \nabla\mathbf{f}$ can be shown in a similar way.

We might expect $f$ to look somewhat like a bell (although I have not checked), meaning that $f''$ would be significantly positive at some large enough $r$. The idea that you could expect the wind to be blowing the opposite way from where it is blowing at your location on a circle of say 300 km around you is very amusing.