Spatial and temporal support of meteorological observations and predictions

ifgi
Institute for Geoinformatics
University of Münster

Edzer Pebesma

DailyMeteo, Belgrade, 24-27 Jun 2014

Overview

1. What is support?
2. Aggregating observations
3. Aggregating predictions
4. An example
5. Prediction errors for aggregations
6. Block kriging
7. Linear vs. non-linear aggregation
8. T-Aggregate-S-interpolate, or S-interpolate-T-aggregate?
9. Bayesian approaches

What is support?

Support is the physical size (in space, in time), of observations, and of predictions.

Why do we need to talk about support?

- we often assume spatial *point support*, but we cannot measure on things of zero size, because measurement devices have a size.
- we often assume instant *temporal support*, but measurement takes time, and sensors need time to respond to changing conditions

Clear support:

- shutter speed of a camera
- filter *time* for PM_{10} measuremet

Not so clear:

- Ground area to which a remote sensing pixel relates (not resolution)
- temporal support of a tipping bucket rain gauge (tip)

What is support?

Support is the physical size (in space, in time), of observations, and of predictions.

Why do we need to talk about support?

- we often assume spatial *point support*, but we cannot measure on things of zero size, because measurement devices have a size.
- we often assume instant *temporal support*, but measurement takes time, and sensors need time to respond to changing conditions

Clear support:

- shutter speed of a camera
- filter *time* for PM_{10} measuremet

Not so clear:

- Ground area to which a remote sensing pixel relates (not resolution)
- temporal support of a tipping bucket rain gauge (tip)

What is support?

Support is the physical size (in space, in time), of observations, and of predictions.

Why do we need to talk about support?

- we often assume spatial *point support*, but we cannot measure on things of zero size, because measurement devices have a size.
- we often assume instant *temporal support*, but measurement takes time, and sensors need time to respond to changing conditions

Clear support:

- shutter speed of a camera
- filter *time* for PM_{10} measuremet

Not so clear:

- Ground area to which a remote sensing pixel relates (not resolution)
- temporal support of a tipping bucket rain gauge (tip)

What is support?

Support is the physical size (in space, in time), of observations, and of predictions.

Why do we need to talk about support?

- we often assume spatial *point support*, but we cannot measure on things of zero size, because measurement devices have a size.
- we often assume instant *temporal support*, but measurement takes time, and sensors need time to respond to changing conditions

Clear support:

- shutter speed of a camera
- filter *time* for PM_{10} measuremet

Not so clear:

- Ground area to which a remote sensing pixel relates (not resolution)
- temporal support of a tipping bucket rain gauge (tip)

What is support?

Support is the physical size (in space, in time), of observations, and of predictions.

Why do we need to talk about support?

- we often assume spatial *point support*, but we cannot measure on things of zero size, because measurement devices have a size.
- we often assume instant *temporal support*, but measurement takes time, and sensors need time to respond to changing conditions

Clear support:

- shutter speed of a camera
- filter *time* for PM_{10} measuremet

Not so clear:

- Ground area to which a remote sensing pixel relates (not resolution)
- temporal support of a tipping bucket rain gauge (tip)

Support of dprec (meteo)

```
> library(spacetime)
> load("st_meteo.RData")
> dprec = m2[NLpol,,"prec"]
> dprec[1,1:2]

           prec
2011-07-01    0
2011-07-02    0

> dprec[1:2,1]

       coordinates prec
1 (4.331667, 51.59)    0
2   (4.433, 52.167)    0
```

?dprec: ... ‘prec’ numeric; daily precipitation amount in mm

What is the spatial support?

Aggregating observations: temporal

Eos, Vol. 93, No. 15, 10 April 2012

FORUM

What Is Monthly Mean Land Surface Air Temperature?

PAGE 156

Land surface air temperature is one of the fundamental variables in weather and climate observations, modeling, and applications. Its monthly mean has been computed as the average of daily maximum and minimum temperatures [*Jones et al.*, 1999]. This is different from the true monthly mean temperature, which is defined as the integral of the continuous temperature measurements in a month divided by the integration period and can be very accurately represented using hourly data, as has long been recognized [e.g., *Brooks*, 1921]. We argue, from scientific, technological, and historical perspectives, that it is time to compute the true monthly mean using hourly data for the national and mean using daily maximum and minimum temperatures.

Technologically, there has been a major shift in the past few decades in temperature measurements, from mechanical thermometers requiring human readings to automated electrical thermometers. These electrical thermometers are in widespread use now because they provide an output signal suitable for use in remote indication, recording, storage, and transmission of temperature data. For instance, the Automated Surface Observing System has been deployed by the National Weather Service in the United States since 1991. Because of the steady decrease in the price of electronic products (including electrical thermometers), even most consumers use digital thermometers in

maximum and minimum temperatures varies from day to day, the difference of the monthly mean maximum and minimum temperatures, i.e., the diurnal temperature range (DTR), is larger than the amplitude of the monthly averaged hourly temperature diurnal cycle. Because regional and global climate change has been presented based on monthly means using daily maximum and minimum temperatures [e.g., *Trenberth et al.*, 2007], it remains to be seen how those conclusions will be quantitatively and qualitatively affected if hourly data are used.

Just like all other meteorological quantities, there could be significant spatial heterogeneities in daily maximum and minimum temperatures (and hence in monthly mean temperature) due to a variety of geographic (e.g., elevation) and transient (e.g., cloud cover) factors. In contrast, transient factors would not affect the true monthly mean using hourly data as much. Therefore, the monthly mean might exaggerate the spatial heterogeneities compared with the true monthly mean.

Because of the possible confusion of modelers about monthly mean versus true

Aggregating observations: temporal

For monthly total rainfall, can we reasonably use

```
> mprec = aggregate(dprec, "1 month", sum, na.rm = TRUE)
```

or should we rather use

```
> mprec = aggregate(dprec, "1 month", sum, na.rm = FALSE)
```

and omit the `NA` values, or should we *estimate*, by

```
> mprec = aggregate(dprec, "1 month", mean, na.rm = TRUE)
> mprec$prec = mprec$prec * 31
```

Counting `NA` values may help deciding:

```
> table(aggregate(dprec, "1 month", function(x) sum(is.na(x)))[[1]])
```

```
  0   1  20  25
333   2   1   1
```

Aggregating observations: spatial

So, `mprec` has monthly total preciptation for Jul 2011 in the Netherlands.
Let us aggregate these spatially:

```
> aggregate(mprec, NLpol, mean)$prec
[1] 131.5018
```

This is the mean monthly total precipitation of:

- **a** the set of 337 rain gauges
- **b** the country (Nederland)
- **c** both
- **d** none of the above

Aggregating observations: spatial - II

```
> aggregate(mprec, NLpol, sum)$prec
[1] 44316.11
```

This is the summed monthly total precipitation for:

a the set of 337 rain gauges

b the country (Nederland)

c both

d none of the above

Aggregating observations: spatial - III

```
> aggregate(mprec, NLpol, mean)$prec
[1] 131.5018
```

This estimates the mean monthly total precipitation of:

- **a** the set of 337 rain gauges
- **b** the country (Nederland)
- **c** both
- **d** none of the above

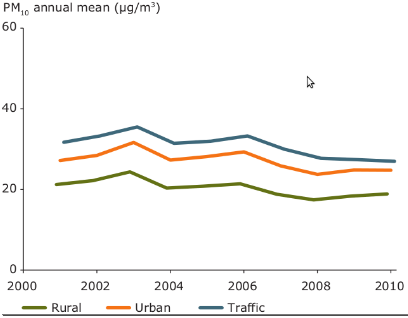
EEA Report | No 4/2012

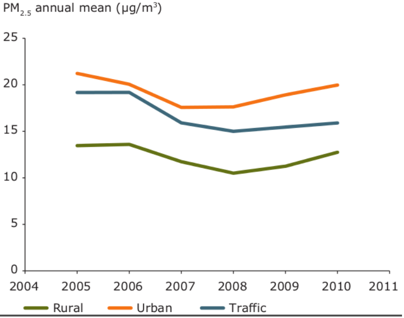
Air quality in Europe — 2012 report

ISSN 1725-9177

European Environment Agency

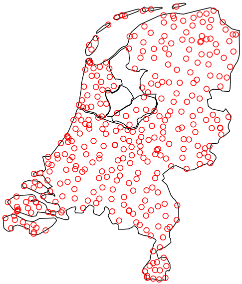
Particulate matter time series, averaged over station type





Spatial distribution of stations over NL

```
> plot(NLpol)
> points(mprec, col = 'red')
```

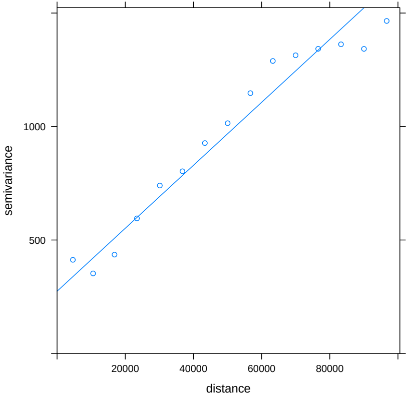


If we want to integrate over NL...

```
> library(sp)
> library(rgdal)
> NLpol = spTransform(NLpol,
+          CRS("+init=epsg:28992"))
> plot(NLpol)
> pts = spsample(NLpol, 500,
+          "regular", offset = c(.5,.5))
> points(pts, col = 'red',
+          pch=3, cex=.3)
```

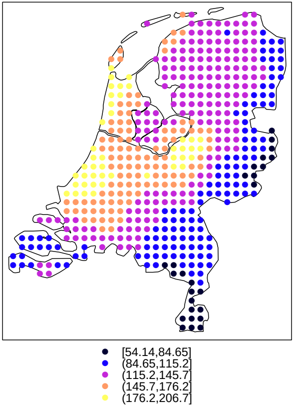
Aggregating predictions

```
> mprec = spTransform(mprec,
+           CRS("+init=epsg:28992"))
> library(gstat)
> v = variogram(prec~1, mprec,
+           cutoff=1e5)
> v.fit = fit.variogram(v,
+           vgm(1, "Lin", 0, 1))
> v.fit

  model       psill range
1   Nug 274.2814774     0
2   Lin   0.0138714     0

> plot(v, v.fit)
```



Aggregating predictions

```
> kr = krige(prec~1, mprec, pts, v.fit)
[using ordinary kriging]
> spplot(kr[1],
+          col.regions = bpy.colors())
```



Aggregating predictions - II

```
> kr = krige(prec~1, mprec, pts, v.fit)
[using ordinary kriging]
> mean(kr$var1.pred)
[1] 129.9698
```

What is the standard error of this mean? Not one of those:

```
> sqrt(var(kr$var1.pred))
[1] 31.02901
> sqrt(var(kr$var1.pred)/length(pts))
[1] 1.391841
> mean(sqrt(kr$var1.var))
[1] 20.30249
```

Prediction errors for aggregations

Block kriging directly estimates

$$Z(B) = \frac{1}{|B|} \int_B Z(s)ds \approx \sum_{i=1, s_i \in B}^{n} Z(s_i)$$

```
> mean(kr$var1.pred)
[1] 129.9698
> krige(prec~1, mprec, NLpol, v.fit)$var1.pred
[using ordinary kriging]
[1] 129.9698
```

Block kriging

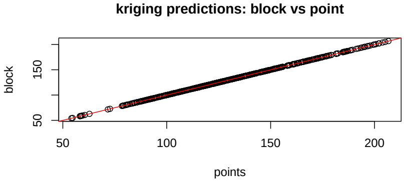
1. Block kriging estimates $Z(B)$, and provides an error measure for $Z(B) - \hat{Z}(B)$, the block kriging standard error.
2. The kriging estimate $\hat{Z}(B)$ is equal to $\sum_{i=1}^{n} \hat{Z}(s_i)$ where the s_i discretize B.
3. The kriging variance (squared standard error) is

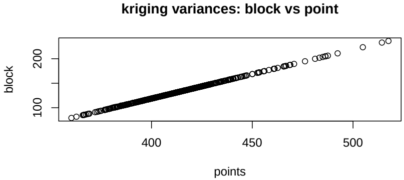
$$E(\sum_{i=1}^{n} \hat{Z}(s_i) - \sum_{i=1}^{n} Z(s_i))^2$$

   and takes all covariances into account.
4. Only the point kriging predictions $\hat{Z}(s_i)$ cannot provide this.

Block kriging: example

```
> p = krige(prec~1, mprec, pts, v.fit)
[using ordinary kriging]
> b = krige(prec~1, mprec, pts, v.fit,
+           block = c(1000,1000))
[using ordinary kriging]
> range(p$var1.pred - b$var1.pred)
[1] -0.10311043  0.03488041
> qqplot(p$var1.pred, b$var1.pred,
+  xlab = 'points', ylab = 'block')
> mean(p$var1.var - b$var1.var)
[1] 281.2661
> qqplot(p$var1.var, b$var1.var,
+  xlab = 'points', ylab = 'block')
```





Linear vs. non-linear aggregation

If, instead of the block mean, we are interested in some other quantity

$$Z(B) = \frac{1}{|B|} \int_B g(Z(s))ds$$

such as the block median or a quantile, how do we get it?

1. not by block kriging
2. we can use *conditional simulation*:
   - draw samples from $(Z|\text{observations})$
   - compute $g(Z)$
   - integrate, spatially
3. this is a (simple) Monte Carlo procedure, not very expensive

Why would you want block kriging?

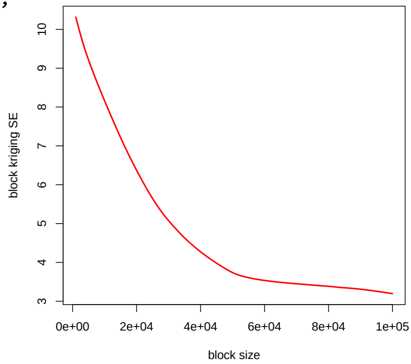
... why would you want point kriging? (think: tipping bucket!)

- when $Z(s) = S(s) + \epsilon(s)$ with $\epsilon(s)$ a white noise process (nugget), block kriging for very small blocks is VERY similar to predicting $S(s)$ instead of $Z(s)$
- ... but block kriging works for any block size, any geometry
- aggregated values, in particular means for larger blocks have lower prediction errors than for smaller blocks
- if $\epsilon(s)$ is, or contains, measurement noise, why predict this?

Block kriging SE as function of block size

For these data, and this variogram model:

```
> p = SpatialPoints(coordinates(NLpol),
+         CRS(proj4string(NLpol)))
> sizes = (1:100)*1000
> f = function(x) {
+         krige(prec~1, mprec,
+         p, v.fit, block =
+         rep(x,2),
+         debug.level=0)$var1.var
+ }
> v = sapply(sizes, f)
> plot(sizes, sqrt(v),
+         xlab = 'block size',
+         ylab = 'block kriging SE')
```



T-Aggregate S-interpolate, or S-interpolate T-aggregate?

ELSEVIER

Geoderma 89 (1999) 47–65

GEODERMA

Spatial aggregation and soil process modelling

Gerard B.M. Heuvelink *, Edzer J. Pebesma

Netherlands Centre for Geo-ecological Research, Faculty of Environmental Sciences, University of Amsterdam, Nieuwe Prinsengracht 130, 1018 VZ Amsterdam, Netherlands

Received 4 November 1997; accepted 6 July 1998

Abstract

Bayesian approaches

- Bayesian approaches take uncertainties in variogram parameters into account;
- the Bayesian framework is comprehensible and consistent, usually prefered
- it may however require supercomputer resources
- in data-rich situations, uncertainty in variogram parameters may not be a meaningful source of uncertainty
- package INLA (`http://www.r-inla.org/` provides nested Laplace approximations to full posteriors, dropping the need for MCMC

Conclusion

The spatial mean of a set of point values in an area is not equal to the mean of the variable over that area!