Disscussion of paper by Carroll et al. for JASA

Peter Guttorp **Wendy Meiring** **Paul D. Sampson**

NRCSE

Technical Report Series

NRCSE-TRS No. 003

Discussion of the paper by Carroll et al.

Peter Guttorp, Wendy Meiring, and Paul D. Sampson*

The problem of trend estimation and surface concentration reconstruction for air pollutants is a difficult one, and we want to start by congratulating the authors for their serious approach to this problem. In this discussion we first describe the physical processes involved in ozone formation. We then mention briefly how to deal with spatial heterogeneity in covariance, argue that it is unlikely that the process is temporally stationary, and finally discuss the authors' choice of space-time correlation function.

1. The chemistry of ozone formation

Tropospheric ozone (O_3) is formed by the recombination of oxygen after ionization by photochemical reactions. It is chemically highly reactive, and an irritant to eyes and respiratory tissues. It occurs naturally in air at about 10 parts per billion (ppb). At concentrations above 100 ppb it is considered unhealthy by the EPA. More than one exceedance of 120 ppb per year yields a non-attainment classification, and an exceedance of 180 ppb is classified as extreme/severe.

The main chemical reactions involved in ozone production (see Baker, 1977, p. 249) are photochemical dissociation of nitrogen dioxide

$$NO_2 + h\nu \rightarrow NO + O,$$

oxidation of nitrogen oxide by ozone

$$NO + O_3 \rightarrow NO_2 + O_2,$$

*Peter Guttorp is Professor of Statistics and Director of the National Center for Environmental Statistics at the University of Washington in Seattle. Wendy Meiring is a Visiting Scientist with the Geophysical Statistics Project at the National Center for Atmospheric Research in Boulder, Colorado. Paul D. Sampson is a Research Associate Professor of Statistics at the University of Washington in Seattle. Part of this research was done with funding from the Electrical Power Research Institute. The NCAR Geophysical Statistics Project is funded by the National Science Foundation.

and ozone production

$$O + O_2 \rightarrow O_3.$$

The higher the concentration of NO_x, the more ozone gets produced. There are many other ozone-producing reactions, several of which involve volatile organic compounds. The NRC report on ozone (National Research Council, 1991) provides more details.

Because of the photochemical character of the above reactions, ozone production is most pronounced during the summer (roughly May-September in the United States). Since a large part of the NO_x-pollution comes from burning of fuels, particularly in automobiles, there is a diurnal pattern to the concentration of nitrogen oxides in the atmosphere. This is further enhanced in some parts of the country by the usual night-time temperature inversion, which in effect puts a lid on the air closest to the ground. In this stagnant night-time air, ozone is rapidly used up in the oxidation process. In early morning the inversion layer breaks up, leading to improved mixing of the air, and a consequent increase in the concentration of NO coming from long-range transport higher up in the atmosphere.

2. The deterministic trend

Like many others who have modeled spatio-temporal data, Carroll et al. use a deterministic trend component and a random process component. Our own modeling of ozone data in California is quite similar (including the decision to model square-root transformed concentrations). We found it necessary, however, to use more complex models for both the trend and the residual random process.

The mean daily summertime ozone concentrations are shown in Figure 1 for three stations in the Sacramento region of California. The concentrations are shown on a square root scale. For these sites there are low concentrations during the night, a rapid increase in the morning hours (corresponding to the break-up of the inversion layer), and a peak in mid-afternoon

(corresponding to peak photochemical activity). One of the stations in Figure 1 (AUB) is in the Sierra foothills at higher elevation than the other two sites. It may at times lie above the inversion layer, whence it does not display as much of a decrease in concentrations during the night. These three sites exhibit the same afternoon peaks.

This discussion has highlighted diurnal patterns in ozone concentration, as well as site-specific differences in these. The authors use a trend which is spatially homogeneous. This implies that there are no remaining spatial trends in the residuals. What is the range of site means of the residuals? Furthermore, this assumption implies that the annual and diurnal patterns are effectively the same across all sites. On the basis of their Figure 6, this may not be unreasonable. There is a further assumption that the diurnal cycle is the same across all months after adjusting for seasonally varying temperature effects.

3. The random process: spatial heterogeneity

The Houston area is geographically homogeneous and flat, so, if we can ignore local emission patterns, it would seem reasonable to assume spatial second-order homogeneity (i.e., a covariance function which depends only on the coordinate difference vector). Carroll et al. also assume isotropy, and do some checking of this assumption. Generally speaking we have not found isotropy a valid assumption for temporally homogeneous data sets. It is possible that the lack of directional covariance structure in the Houston area data can be explained by the averaging over entire years, with attendant masking of seasonally dominant effects due to changing meteorological conditions. Such effects are to be expected because of the photochemical nature of ozone production, as described above. Additionally, it was unclear to us what the units of distance were. There is some indication that distance was simply computed in “degrees”, without accounting for the difference between latitude and longitude in computing geographic distances. Their fitted spatial model may therefore be (unintentionally) anisotropic.

When dealing with ozone for areas with considerable orographic structure it is not reasonable to assume spatial homogeneity. Since monitoring data typically contain reasonable numbers of replications (albeit temporally dependent) it is possible to estimate even heterogeneous spatial covariances (Guttorp and Sampson, 1994, present a variety of methods for doing this).

To illustrate how we deal with spatial heterogeneity, we consider some ozone data from the Central Valley of California. The data were collected as part of the SARMAP project, a multi-agency effort to develop and assess a deterministic regional tropospheric ozone model. Figure 2 shows the ozone monitoring sites. Ozone data were collected hourly for two months in the summer of 1990 at 131 sites, of which we use some 100 sites for our covariance estimation (the remainder of the sites were reserved as a validation sample). Our approach is to take out the mean structure on the square root scale (as shown in Figure 1), prewhiten the residuals using AR(2)-models for each site, and then to compute correlations of the residuals from this model. We then transform the geographic map to a new map in which residual correlations are approximately isotropic. Since the entire convex hull of the geographic coordinates is transformed, we can estimate the covariance between unobserved sites by measuring their distance in the transformed plane (for details, see Guttorp et al., 1994, and Guttorp and Sampson, 1994). Figure 3 illustrates the transformation for the 2 pm hour. We see that sites in the southern central part of the San Joaquin Valley are pushed closer together, and (relatively speaking) sites in the San Francisco Bay area are pushed away from the more southerly sites. In other words, there is substantial intersite correlation in the south central valley, and these sites show relatively low correlation with the more northerly sites.

4. The random process: temporal stationarity

The model by Carroll et al. is temporally stationary. In our studies of California ozone we have found strong within-day structure. Figure 4 depicts the correlation structure, in terms of

variograms (labelled “Dispersion”) in the transformed plane, for a subset of the SARMAP residuals, corresponding to 32 stations in the Sacramento region (shown highlighted in Figure 2). As seen in Figure 4, during the early morning there is very low correlation (high dispersion) between sites, and a high nugget effect, while the early afternoon ozone readings exhibit stronger correlation even at substantial distances. Consequently, we have found that the spatial correlation structure depends strongly on the hour of the day, at least for short temporal lags. Since this is consistent with the mechanism of ozone production described in section 1, we find the temporal stationarity assumption rather dubious.

5. The space-time correlation model

Our analyses of ozone over a similar spatial domain suggest spatial and temporal heterogeneity in the spatial correlations at this scale. An assumption of spatial homogeneity is probably more reasonable for the Houston region, but the assumption of temporal stationarity is highly suspect. Our analyses have shown that both the temporal and spatial correlation structure of tropospheric ozone vary with the hour of the day, being substantially different in the night/early morning hours and in the mid/late afternoon hours. Fitting a stationary model to nonstationary data may explain some of the unusual features of their fitted model.

We note first that three of the seven correlation parameters in their model are at, or nearly at, the boundary of the parameter space (ψ_1, ϕ_2, and ϕ_3 are all approximately equal to 1). What does this say about the model? Second, the pure temporal correlations (i.e., correlations for distances $d = 0$) are 1.0, 1.0, 0.69, 0.01 for temporal lags 0 to 3. The 1-hr lagged correlation of 1.0 would represent an unusual time series. For $d > 0.3$ we observe a surprising correlation structure with high 0-hr (contemporaneous) and 2-hr lagged correlations, but essentially zero 1-hr lagged correlations. (The authors try to explain this, but it still strikes us as unusual.) And, finally, it is surprising that there is no spatial decay in the 2-hr lagged correlations. These features, our doubts

about temporal stationarity, and the doubts About the validity of the correlation model, leave us unconvinced that the proposed space-time correlation is, in fact, suitable for modeling hourly tropospheric ozone.

References

Baker, B. G. (1977): Control of noxious emissions from internal combustion engines. In J. O'M. Bockris (ed.): *Environmental Chemistry*, Ch. 9, pp. 243—284.

Guttorp, P., and Sampson, P. D. (1994): Methods for estimating heterogeneous spatial covariance functions with environmental applications. In G. P. Patil and C. R. Rao (eds.): *Handbook of Statistics XII: Environmental Statistics*, Elsevier/North Holland, New York, pp. 663—690.

Guttorp, P., Meiring, W., and Sampson, P. D. (1994): A space-time analysis of ground-level ozone data. *Environmetrics* **5:** 241—254.

National Research Council (1991): *Rethinking the ozone problem in urban and regional air pollution*. National Academy Press, Washington, D.C.

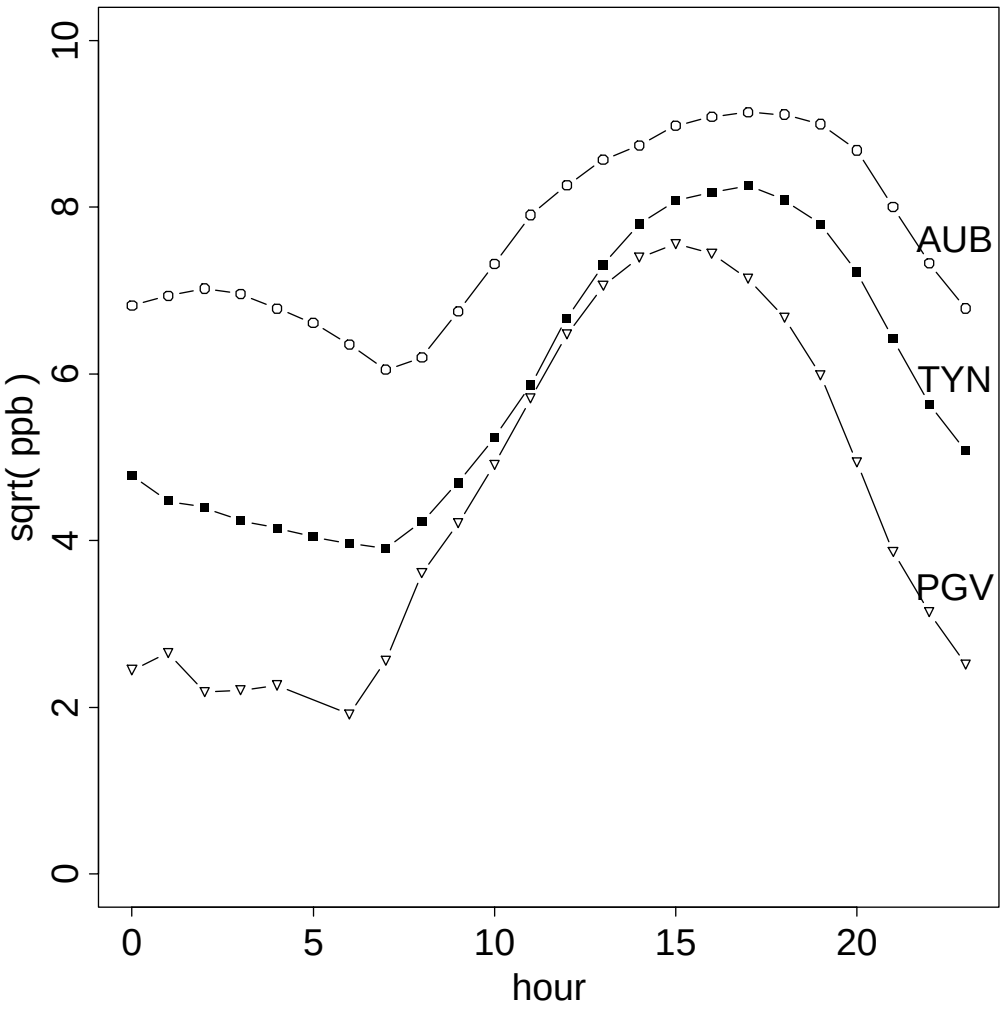
Figure captions

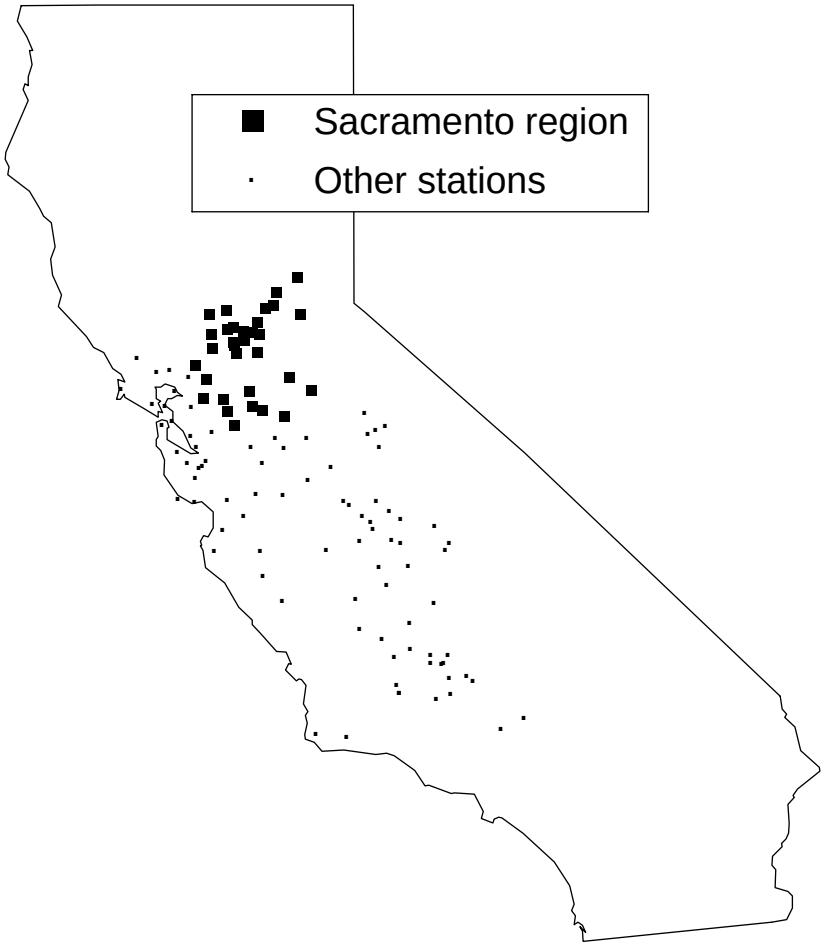
Figure 1. Hourly mean ozone levels on the square root scale for three surface ozone sites: Auburn—Dewitt—C Ave (AUB) at elevation 430 m, Pleasant Grove (PGV) at 8 m, and Tyndall—Road 97 at 9 m.

Figure 2. Map of SARMAP surface ozone sites. Sites in the Sacramento subregion are highlighted.

Figure 3. Geographic map (left; coordinates in km) and deformation map (right) of the 2 pm covariance structure for residuals in the San Joaquin Valley of California.

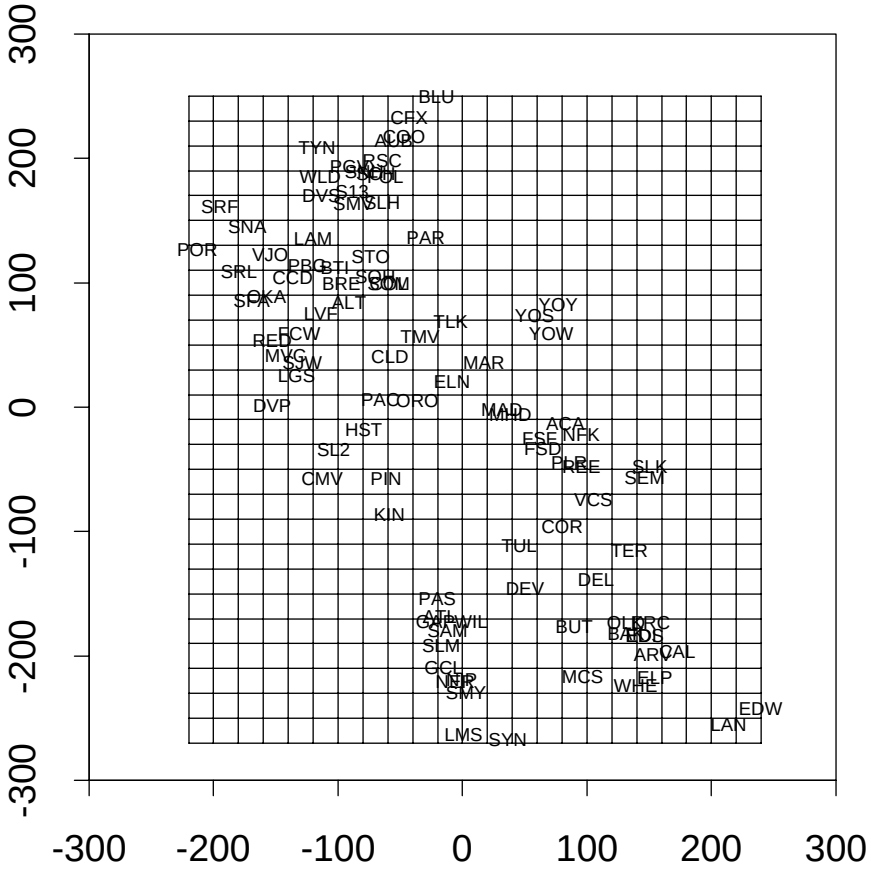
Figure 4. Deformation maps and fitted variograms based on 32 sites in the greater Sacramento region, calculated based on the pre-whitened residuals for different times of the day. Each plot is based on 3 hours (centered at the hour indicated).


sqrt(ppb)
0
2
4
6
8
10
0
5
10
15
20
hour
AUB
TYN
PGV

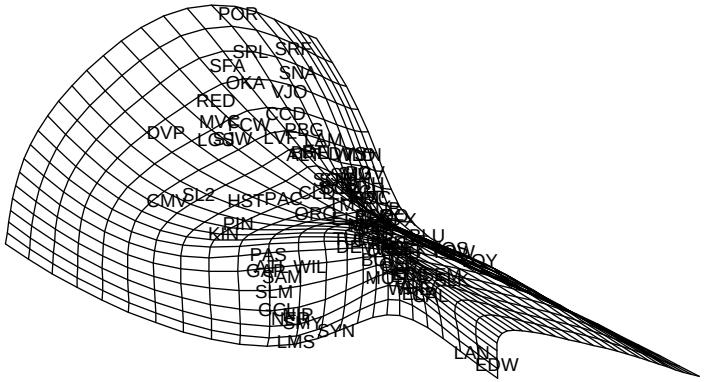

Sacramento region
Other stations

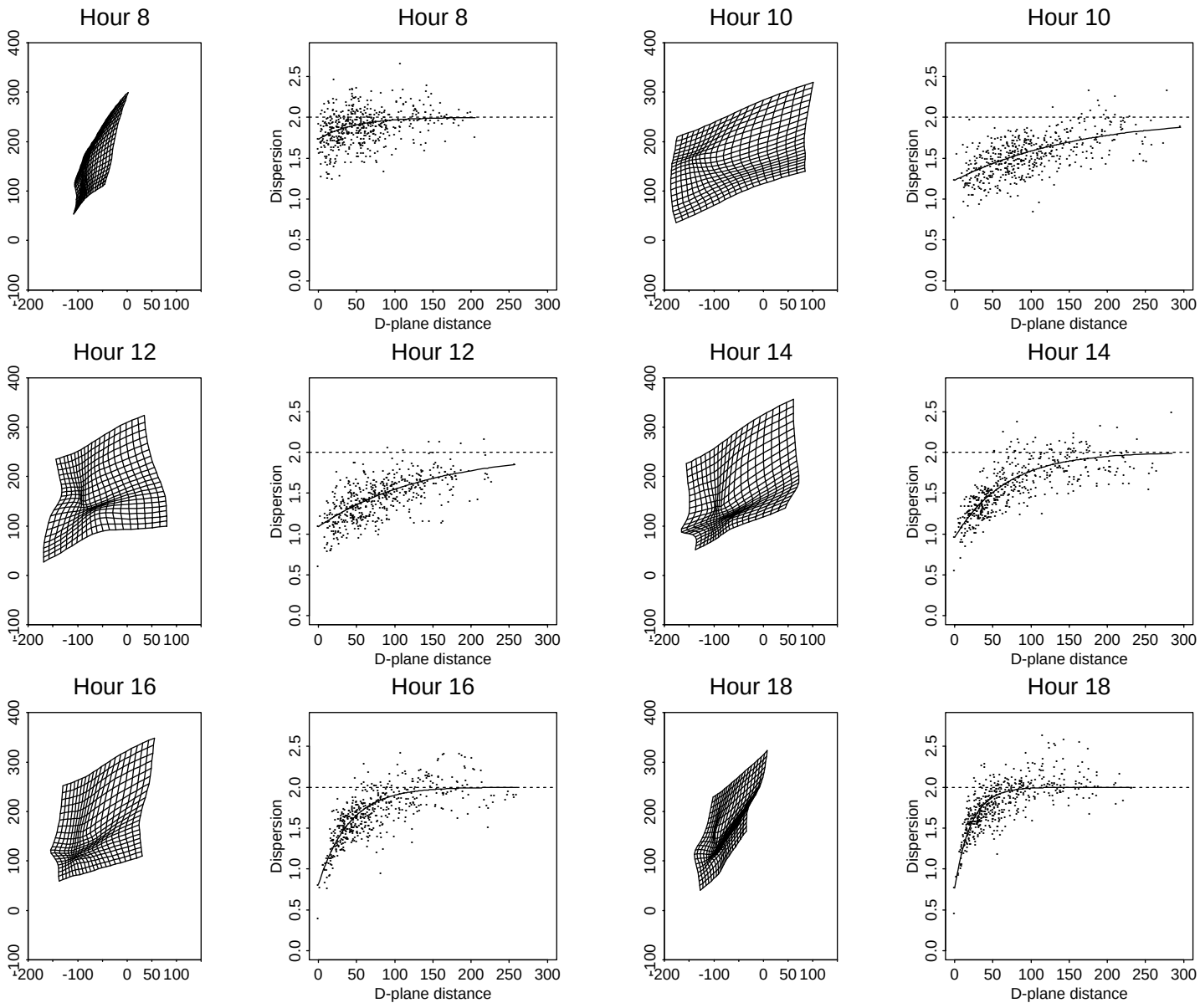
Hour = 14

G-plane Coordinates



D-plane Coordinates




Hour 8
Hour 8
Hour 10
Hour 10
Hour 12
Hour 12
Hour 14
Hour 14
Hour 16
Hour 16
Hour 18
Hour 18
Dispersion
D-plane distance