ISSN 1946-7664. MCFNS 2023
Vol. 15, Issue 1, pp. 1–7
Mathematical and Computational Forestry & Natural-Resource Sciences

AVAILABLE ONLINE AT HTTP://MCFNS.COM

Submitted: Dec. 6, 2022
Accepted: Mar. 14, 2023
Published: Apr. 30, 2023
Last Correction: May 2, 2023

NEAREST-TREE AND VARIABLE POLYGON SAMPLING

Kim Iles
Kim Iles & Associates Ltd.
412 Valley Place, Nanaimo, BC, V9R 6A6, Canada

Abstract. Sampling a nearest neighbor is often presented as a Hansen-Hurwitz or Horvitz-Thompson estimation process. This may not be the most informative viewpoint, and measuring the probability of selection is not necessary. The measurement of the nearest object as a "depth" over the selection area can be done by a sampling process, and the total estimated without knowing the polygon areas. The process is unbiased, quite general, and easy to understand. It can be extended to more than just the nearest object to a sample point and to many different polygon shapes. This paper is an extension, simplification and generalization of an earlier paper in this journal (Iles, K. 2009. "*Nearest-tree" estimations—A discussion of their geometry*, MCFNS 1(2), pp. 47–51.), but does not require a random orientation or weighting for the direction of measurement from the tree to the polygon border.

Keywords: n-tree sampling; nearest tree; Voronoi polygons; unbiased estimates.

1 The Discussion

Most of forest sampling is about geometry, either plane or solid geometry. Typically, the area on which a single tree is selected gives a probability of sampling, often put into the familiar Hansen-Hurwitz equation as a statistical approach. The nearest-tree approach is an example of selection proportional to a Voronoi polygon around the tree. Forest samplers have learned long ago that the logic of sampling is better viewed as a plot surrounding a tree, rather than a tree falling inside a plot. The chance of selection is the area of the polygon around the tree divided by the total area.

There is a well-known and often cited publication by Overton (1995), but the point is often missed that he placed his paper in the "Teachers Corner" section of that journal. His intention was to emphasize the generality of the Hansen-Hurwitz viewpoint in statistics, where the estimate of a population total from a single object is always:

$$Total = \frac{Object\ measurement}{probability\ of\ object\ selection}.$$

I will use volume as an example, but the same process applies to value, weight, carbon, or any other tree (or object) measurement. If the random point in the area falls into a polygon covering 1% of total tract area, the total volume as a Hansen-Hurwitz expression is:

$$Total\ volume = (tree\ volume/1\%) \qquad (1)$$

or

$$Total\,volume = tree\,volume \times \left[\frac{total\,area}{tree\,polygon\,area}\right] \qquad (2)$$

In Equation (2) above, which is more intuitive but less general, the term in square brackets is often called the "Expansion Factor". If several objects are chosen at a sample point, the estimates for each of the objects can be combined, as in the Horvitz-Thompson approach for a cluster of sampled objects with different selection probabilities. The problem in practical terms is to make any measurements involved with a reasonable level of effort. The tree volume is straightforward, but getting the polygon area is not trivial in the field, except for those plots that are circular or of fixed area. This difficulty can be avoided.

I would like to suggest the following view of the situation. An area has polygons upon it. In the simplest case, the polygons tesselate the area, so the total polygon area is therefore known beforehand. If not, a simple count of the polygons selected will determine the relative area of their coverage. An example of this might be counting tree crowns on an aerial photo. If the average count of random or systematic points that fall into one or more polygons is 0.73 polygons per point, it means that the total area of the polygons could be estimated as 73% of the total area. If the polygons overlap, giving an average count of 3.8 on the sample points, the estimation of the total of polygon areas is 380% of the total

surface area. The shape and size of the polygons are of no consequence in this estimate. Foresters perform such counts frequently with both fixed and variable plot sampling, and less frequently with irregular polygons - such as tree crowns or vegetation patches.

In the simplest case, the total of polygon areas is perfectly known as the area of the tract of land involved. For the moment, let us assume that this is the case. What we really want, of course is the total of some property of the trees, with volume as one example. Visualize the volume of the tree as a solid made of ice, perhaps one with the same shape as the tree itself. Now let it melt, and fill the area of each polygon. The polygons have different sizes and tree volumes, and therefore have different depths to which their polygons would be "filled" across the land area. It is obvious that the total volume of the tract is simply the average depth of the water times the area of the tract, much like calculating the volume of a lake, the volume of a pile of wood chips, or an ore body sampled with vertical probes. We simply have to determine this average "depth" by sampling.

The area of the polygon is measurable, but not trivial. It may be easier if the border is made of straight lines between known points, as land surveyors do all the time. GIS systems can do this easily as well. The GIS can determine the polygons, sum them, select some to sample, and help you find those polygons in the field. These polygons do not need to be "correct" in some way, you simply need the polygon areas - and perhaps some border locations on the ground during the field measurement phase.

In cases without computer assistance, you will choose a random point, and need the area for the polygon involved with that decision. The polygon area is difficult to measure in the field, but we do not need to do that. We need the average volume/polygon area, and we can sample for that.

The polygons in Figure 1 surround trees, and also provide a fixed reference point inside the polygon. We know which Voronoi polygon we are in, based on the closest tree to a random point. Imagine that the polygon around the tree is now entirely divided into walled triangular "gutters" that will collect the tree volume when it melts. Now we have a population of vanishingly thin triangular "gutters" that will fill to different depths, with sides of different radii "R_g" in each case. The average of all those depths times the total tract area will now provide the average volume of the tract. The gutter areas are chosen proportional to their areas by the random point selected for sampling, so a simple average can be taken. A useful view of the situation is to simply concentrate on the triangles, and ignore the fact that they happen to also combine to form Voronoi polygons around trees.

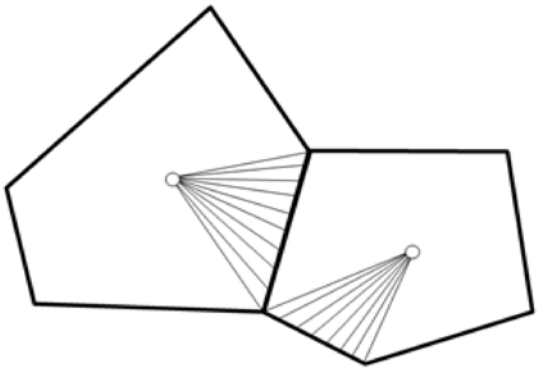

Figure 1: Polygons divided into "gutters".

You are as sure of the total volume as you are of the average depth inside the triangles, and you can sample for this value. The statistics are trivial. How deep is the horizontal component inside each triangle? That is a simple computation:

$$Depth = \frac{Tree\ Volume}{\pi R_g^2} \tag{3}$$

To determine the distance R_g, simply move outward from the sample point, along a line connected to the tree, and find where the distance to the sample tree is equal to the distance to the next closest tree along that line. This finds the boundary that is a bisector between the trees and is on the border of the Voronoi polygon around the sampled tree. If this proves inexact in the field, you can also measure 3 distances to calculate the distance to the bisector forming the polygon edge, as in Figure 2 from an EXCEL worksheet using standard algebra for the known sides of a triangle. The author would be happy to supply a copy of this spreadsheet. It is also possible to get the R_g distances approximately in most cases, then correct them by a more exact estimate on a subsample basis, as we often do in forest sampling. The distance desired in the example is 20.85m. The distance is not the minimum to the polygon edge, but the length to the border along the line connecting the sample point and the nearest tree. An equivalent process is the measurement we have traditionally done in Variable Plot Sampling. Tree volumes, or other values, are divided by tree basal area (or more generally, by tree plot area). This form of "depth" on the sample base is typically called the "VBAR", or "Volume to Basal Area Ratio". In that case, of course, the total area of tree circles inside the tract is an estimate as well, and the statistics are treated as a double sample, or frequently as a product sample where the estimated total tree basal area is multiplied by the estimated average VBAR. The tree polygon area and the average VBAR can be esti-

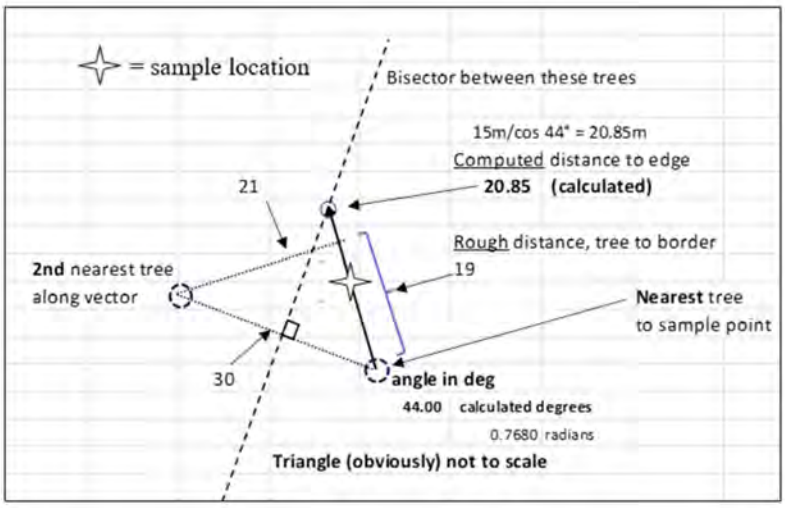

Figure 2: Calculating the distance from a tree to the polygon edge.

mated by separate samples, but are usually done at the same sample points, and the correlation is virtually zero with any stratification based on tree heights. The VBAR can be subsampled with only some of the trees at each sample point, of course. In essence, this is the case with gutter subsamples as well.

The units of the depth ratio would be something like "cubic metres per square metre of tree polygon area". Any other measurement is handled the same way, and might be dollars or tons per square metre. Numbers of trees are handled simply by dividing "one tree" as a measurement by the area involved for the polygon around the tree – giving "trees/square metre" for the total polygon area of the tract. Species and other nominal variables are handled the same way.

Ecologists have long attempted to estimate tree numbers from nearest tree distances, adding some additional distance because they needed something approaching the distance R_g, as noted above, but did not seem to have noticed this simple way of obtaining the correct distance; or perhaps I have simply been unable to find such a reference. I believe that the reason this eluded many of them was that they typically visualize the distance around a sample point as a circular plot – rather than approximating the area of the polygon around the tree, although they sometimes did mention the selection area of the tree as a Voronoi polygon. Assuming random distributions to avoid the previous sampling biases or to adjust the distance measurements seems a particularly dismal way to approach a solution for the bias, and more complicated distributions have done little to solve the problem either.

An unbiased solution for estimates selected from the nearest tree is just as simple as getting the distance to the polygon boundary. This generalizes to sampling using the distance to the boundary of any polygon selecting the object for sampling, not just Voronoi polygons. Fixed and Variable plots are a special case of this situation. One might consider this process to be "Variable Polygon Sampling" for any measurement inside a polygon of arbitrary shape. Simple shapes, not involving "gaps" along a line from a reference point, are much preferable in this approach, but not absolutely required. How the polygon is constructed is of no interest in terms of unbiasedness, though it might affect the efficiency or ease of sampling.

Publications such as Fraser (1972) chose to connect the measured variable to a single triangle connected to each tree, presumably because it was easier in the field to measure the area of that triangle than a Voronoi polygon. I would agree. My own view, however, is that this is less flexible in weighting the triangles by various means, and I have found the triangles a bit difficult to reliably locate and associate with the correct tree in the field. The geometric viewpoint of the measured variable as a "depth" over these triangles is virtually the same.

2 Polygons by computer applications

There is an alternative to determining polygon area in the field. With automated methods of scanning forest stands, the polygons surrounding trees can be automatically generated. Software is currently able to do this, and the key point is not that it involves LiDAR, but the fact that it is done automatically by computer algorithms. Other forms of remote sensing might be more efficient. The fact that you can deliver the total of polygon areas, choose the samples, and specify sample polygon

areas reduces the field effort. Sampling can be done randomly, or perhaps much more efficiently as a systematic sample. The only additional step needed is to get the tree measurement involved – requiring that the location can actually be found on the ground. A more efficient method might be to choose clusters of polygons, especially if these are related to easily located outer polygon boundaries on the ground like roads or openings. It is not apparent to me that tree polygons can currently be reliably located on the ground – other than sparce trees in some forest types. It is, however, apparent to me that it will be possible to do in the future.

Initial rough estimates of the results would also allow more efficient systematic sampling, and all these benefits are available before going to the field. The GIS system can total the polygon areas, so the entire area need not be tessellated. Unproductive ground can be stratified out of the process, and better polygon areas might be created with alternative algorithms and estimated tree values. Any rough tree volume estimates do not need to be unbiased, since the estimate will be corrected by ground sampling. The book by Okabe (1992, page 128) discusses alternative computer methods for polygon constructions using weighting. This approach does not assume that there is only a single tree in these polygons – just that whatever is inside the polygon can be identified and measured. Sometimes you can adequately assign simple variables without field work.

In some instances, the volume inside the polygon might be zero. A characteristic that might be measured in the polygon is assumed to be associated with the "central point", which would normally be the center of a tree stem – but not necessarily. You could use 10cm from the North side of each tree bole, for instance, which could be more precisely located on the ground. The total polygon area would also include undetected vertical stems or fallen trees. The computer can also specify the approximate distance to the edge of the polygons from the sample tree in various directions to facilitate work in the field, but that would seldom be necessary.

3 Sampling Tree polygons that overlap

Suppose we want to measure the values of the 3 closest trees to any sample point. Around the border of the nearest (1^{st}) tree polygon there is a string of small polygons (which I would call "shards") that surround that initial polygon and identify that tree as the 2^{nd} nearest tree. This forms a larger outer boundary for a continuous polygon where that tree is either the 1^{st} or 2^{nd} tree. Likewise, an additional "necklace of shards" can be added to that border to identify the polygon area where the tree is the 1^{st}, 2^{nd} or 3^{d} tree that would be selected by random sample points, as shown in Figure 3.

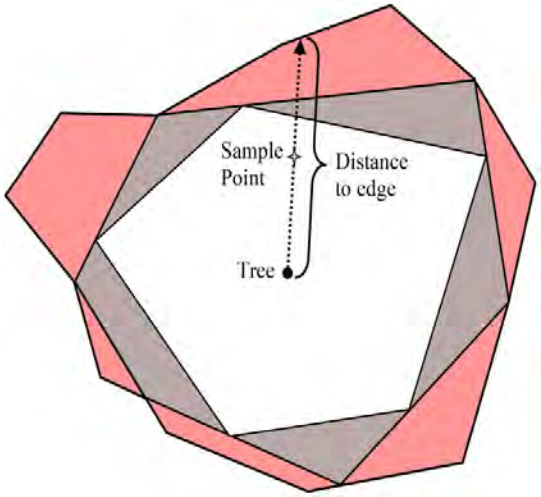

Figure 3: Finding the perimeter of the 3^{d} level polygon.

The surrounding shard necklaces of individual trees on the tract intersect (at the nodes) without overlapping (Figure 4), and they tesselate the area. They are

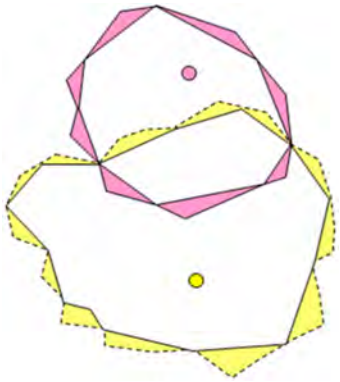

Figure 4: intersecting shard necklaces.

interesting geometrically. They might also provide an original type of jigsaw puzzle or stained-glass window, but in this example they just help to specify the 3 trees involved, and determine the outside boundary of each overlapping tree polygon. This graphic is only an example for illustration, rather than a correct rendering. I believe that Okabe might refer to these as "order-3 Voronoi polygons" (page 154, and well illustrated in that text).

At any sample point, 3 polygons overlap, as shown in Figure 5, and we would need to estimate the Volume/area relationship for each of them (at their full size). We therefore need the distance to the edge of each of the 3 tree polygons. These are the border points we want to identify. We simply retreat along a line through

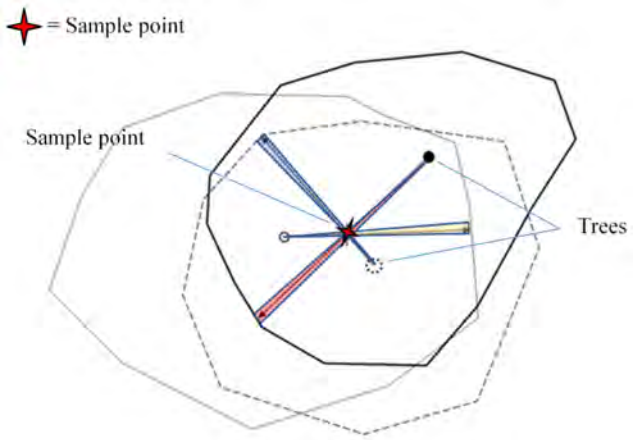

Figure 5: Overlapping polygons that choose 3 trees.

the random sample point to each of the trees, until we reach the border where each tree would change to the 4^{th} most distant tree. If the polygon areas have already been determined by a computer, we only need to locate and measure the items in the polygons of known area, and we have a single average "depth" for each polygon.

These polygon necklaces completely tesselate the land base at each level. Since every point has one and only one second-nearest tree, for instance, it is clear that the second level shards tessellate the land base, and therefore have the same combined area as the central polygons that they surround (since they have the same total area and number of trees). They do not, however, individually have the same area as their surrounded center polygon. The combined areas of the center polygons plus their 2^{nd} level shard necklaces obviously overlap and form a total area 2X the land base. The same holds true for other necklaces of shards, so the 3^d level polygons overlap and cover the area 3 times, or 300% of the land base.

Distance from tree to each polygon border is shown.

This approach is one way to choose trees in clusters, and for which we can apply the Horvitz-Thompson approach to the trees, each of which has a polygon where it would be one of 3 trees closest to a particular sample point. For each tree (if the computer does not provide that area) we deduce the depth along the gutter reaching out from each tree. In this case, there will be 3 trees with gutters that overlap each other as they cross the sample point. Adding the 3 depths will produce the estimate that is needed. The sum of all 3 depths times the area of the tract gives an estimate of total volume. The same number is produced by the average gutter depth times 3X the tract area. This is virtually what is done with Variable Plot Sampling, when the estimating equation uses the average of the VBAR measurements for a variable number of trees that are "in" with the angle gauge. In this application we always have 3 trees, and the advantage is that we know the sum of all the areas involved, where in Variable Plot Sampling the total basal area is only estimated. Subsampling is also obviously possible.

You could also use the simple Voronoi polygons, and expand their dimensions to form a variable number of overlapping polygons. This is simpler to do by computer, and the process would give automatic tree selection at sample points with known polygon areas. The process is very general.

It has not escaped the author that this process might be extended to higher dimensions and nearest-neighbor methods, but that is beyond the scope of this current paper. It is also worth mentioning, that when more than one polygon is involved on a plane, it might be an advantage to use distance-variable estimators (Iles, 2007), rather than a simple "flat" depth measurement. This might be visualized as the tree melting as wax might, giving different depths inside the polygon or gutter. The depth estimates might then compensate to some extent whenever objects avoid one another, as they often do in nature. For repeated measurements such as permanent sample points, this approach is more than 40 years old, and has been used to lower variability and produce "consistency" over time for surviving trees to provide less variable forest growth estimates (Iles, 1979 and 1989; Flewelling, 1981; Therien, 2011; McTague, 2013), but sudden changes due to mortality have remained a problem.

The realignment of polygon size and shape from mortality will mitigate the affect of mortality on live tree volume caused when trees are suddenly removed from the population of live trees. With mortality, a new live tree and its adjusted polygon will replace the one that died, and the bordering tree polygons would also adjust themselves. Volume growth over the point would often still change slowly for permanent variable plots on the polygons of surviving trees, although a bit more than previous methods. When dying trees are replaced by another surviving tree and polygon, the total live volume and growth result should be more stable. Simulations with actual stand maps would provide an insight into this. My own view is that the process would be better done with a set of systematic points in a line. This would allow the maximum number of readjustments on tree polygons when relatively sparse mortality occurs, because of the larger number of bordering polygons for the points along the line.

4 Extensions to k^{th}-nearest tree methods

I will use a 6-tree cluster as an example. In this case, the 6 nearest trees to a sample point are chosen. You estimate the R_g distance of only the 6^{th} nearest tree polygon. That tree is chosen when you fall into any of the 6^{th} level necklace of shards surrounding the tree. All 6 trees are now added for volume, and divided into that single polygon area. This is usually presented as a fixed plot using the distance to the 6^{th} tree as a kind of plot radius. Since you are inside the 6^{th} shard, you are already close to the border distance needed. Historically, a small distance is typically added to the initial distance from the sample plot to the 6^{th} tree, and that distance should be the distance from where the sample point fell in the shard to the outer edge of the polygon bordering the 6^{th} tree. In this method, only the distance to the 6^{th} tree is used. Even when the correct distance R_g is used, it is clear from counter-examples that this method using only the polygon of the 6^{th} tree would be biased, and simulations seem to verify this. On the other hand, if the bias is shown to be small in widely consistent actual forest types, should we care that much? The process is less difficult than specifying all the polygon estimates and distances, although it is prone to edge-effect biases.

5 Practical applications

Nearest-tree, or n-nearest trees approaches are not a problem because of biases. One could specify a handful of unbiased methods easily, or perhaps show that any uncorrected bias was of no practical consequence, but that is not the issue. The real question is whether the method is a good alternative in comparison to other available methods. Is it better than stopping at sample points, quickly counting the trees with an angle gauge, perhaps measuring one or two trees and moving on? If not, how much should we dwell on a less efficient method of sampling? Are there situations where automation would relieve the mechanical issues of selection and measurement in specific situations and make this process desirable?

Clever field application of geometry ideas such as tree selection with angle gauges or clever probability determinations such as with Randomized Branch Sampling and Adaptive Cluster Sampling have often been the key to practical advances. Certainly, these can all be reformulated as Hansen-Hurwitz or Horvitz-Thompson algebra – but the application of them has been made practical by previously unimagined field procedures. Perhaps this viewpoint of the process of sampling inside a chosen polygon might also be valuable.

Acknowledgement

The author wishes to thank anonymous reviewers for their suggestions.

References

Boots, B., K. Sugihara, S.N. Chiu, and A. Okabe. 2009. Spatial tessellations: concepts and applications of Voronoi diagrams. 678 p.

Flewelling, J.W. 1981. Compatible estimates of basal area and basal area growth from remeasured point samples. Forest Science. 27:191–203.

Fraser, A. R., and P. Van den Driessche. 1972. Triangles, density, and pattern in point populations. In: Proc. of the Third Conf. Advisory Group of For. Stat., Int. Union for. Res. Organ., Inst. Nat. Res. Agric., Jouy-en-Josas, France (pp. 277–286).

Iles, K. 1979. Some techniques to generalize the use of variable plot and line intersect sampling. Forest Resource Inventories, workshop proceedings (W.E. Frayer, ed.) v1, p. 270–278. Colorado State University, Fort Collins. 2 vol., 1037 p.

Iles, K. 1990. Critical height sampling: a workshop on the current state of the technique. Iles, K., 1990. Critical height sampling: a workshop on the current state of the technique. USDA Forest Service general technical report PNW-GTR-Pacific Northwest Research Station (USA). ISSN: 0887-4840. Pp: 74–83.

Iles, K. 2009. "Nearest-tree" estimations—A discussion of their geometry. Mathematical and Computational Forestry & Natural-Resource Sciences (MCFNS), 1(2), pp: 47–51. Retrieved from `http://mcfns.net/index.php/Journal/article/view/MCFNS.1-47`.

Iles, K. 2010. Nearest Neighbor Bias—A simple example. Mathematical and Computational Forestry & Natural-Resource Sciences (MCFNS), 2(1), pp: 18–19. Retrieved from `http://mcfns.net/index.php/Journal/article/view/MCFNS.2-18`.

Iles, K., and H. Carter. 2007. "Distance-variable" estimators for sampling and change measurement, Canadian Journal of Forest Research 37(9):1669–1674.

McTague, J.P. 2013. The use of composite estimators for estimating forest biomass and growth from permanent sample plots established by the angle count method. In Proceedings of the 2013 Joint Statistical Meetings, Montréal, QC, Canada (pp. 3–8).

Overton, S., and S. Stehman. 1995. The Horvitz-Thompson Theorem as a Unifying Perspective for Probability Sampling: With Examples from Natural Resource Sampling. https://www.tandfonline.com/journals/utas20, Vol. 49, https://www.tandfonline.com/toc/utas20/49/3.

Therien, G. 2011. Relative Efficiency of Point Sampling Change Estimators. Mathematical and Computational Forestry & Natural-Resource Sciences (MCFNS), 3(2), pp: 64–72. Retrieved from http://mcfns.net/index.php/Journal/article/view/MCFNS.3-64.