Colorado River Stochastic Database Development
Project Task 1 – Enhancement of existing database

Draft Report on Current Status of Database
Jeffrey D. Rieker – USBR TSC
September, 2004

1.0 - Background/Introduction

The purpose of this document is to summarize the current dataset upon which the stochastic analysis will be based, and report on any problematic issues with the existing dataset as well as suggested methods to rectify those issues. This report represents the product of Task 1.1, and suggests methods to be used in Tasks 1.2 and 1.3.

2.0 - Summary of Data

The existing dataset used for this analysis was obtained from Jim Prairie of the USBR's Upper Colorado Regional Office (UC office) on September 9, 2004. An alternative dataset was provided by Dr. Jose Salas from Colorado State University (CSU) on September 18, 2004. The datasets were generally identical from the beginning of the record (10/1905) through January of 1971. There were large differences between the datasets for the period starting January, 1971, and ending with the end of the CSU record in September, 1990. Verbal communication with Jim Prairie indicated that all data after January, 1971 is considered provisional, and has been altered in an attempt to improve accuracy of the data. Since the UC office data is considered to be the most current existing data, and currently extends through 1995, this dataset was used for the analysis.

The data consists of 29 monthly timeseries records, representing localized inflows to sites throughout the Colorado River basin. The localized inflows were accumulated to produce total naturalized flows at each site. This analysis included a verification of the accumulated values. The existing dataset begins in October, 1905, and extends to December, 1995. An extension of the data through the year 2000 is expected this fall.

This analysis included a check of the total natural flow data provided in both datasets, to verify that the total was equal to the sum of the local flows. The following conclusions were reached as a result of this verification.

1) The calculations of the total natural flows based on the localized inflows match those provided by the UC office (with the exception of the first three months, which were apparently in error due to the erroneous localized flows on site 24 in the provided dataset).
2) The calculations of the total natural flows based on the UC data match the CSU calculations up to 1971, with the exception of sites 20 and 28 because of the differences in node topology (noted in the following section).

3.0 - Summary of Data Issues, and Suggested Enhancements

A number of issues were identified with the existing data that may need to be rectified prior to the commencement of the stochastic analysis. The issues identified with the current dataset are outlined in this section. Additionally, the suggested enhancement to rectify the issue is located immediately below the issue description in alternate font.

3.1 – General Issues

1) The map shown in Figure 1. of the *Statement of Work* appears to be missing site 20.

   The map should be improved to include site 20 by the people who originally generated the map. Ideally, a new map should be generated to better reflect the locations of each of the subbasins.

2) It needs to be identified whether the Paria River should be included in the Lee's Ferry total natural flow (site 21 in site 20… currently my calculations and the UC calculations do not include it, however the CSU database calculations did), and whether the Bill Williams River should be included in the Parker total natural flow (site 27 in 28… similar discrepancy between UC and CSU).

   The topology set forth by the UC calculations will be used.

3) The data obtained from the UC office matches the data obtained from CSU up to 1971, where they differ due to newer data that is being used by UC/LC.

   No action is necessary, the data provided by the UC office will take precedence over the CSU data.

4) According to previous documentation, the historic data (prior to 1971) found on microfiche does not consistently correspond to the 'official' data that is currently being used. From speaking with Jim Prairie, a previous rapid analysis showed that most sites did not differ appreciably, but the San Rafael microfiche data was apparently very different than the 'official' dataset. These discrepancies are scheduled to be investigated at a later date. I did a rapid spot check, and verified that most of the 19 records were accurate. Site 3 (0910900) was missing from the microfiche transmittal, and site 19 (0937950) contained at least one deviating data point (10/1907), which illustrated that the existing dataset may have minor data deviations. Additionally, the San Rafael site (17 – 0931500) was verified to be very different from the data contained in the 'official' historic record that was provided.

   Staff time should be used to check all of the records against the microfiche records, and rectify any major deviations. Additionally, a decision needs to be made on whether to use the existing, 'official' record for San Rafael, or the microfiche record.

3.2 – Short Records

5) Visual inspection of the data revealed that average monthly localized flows were found in following lower basin sites and years:
   a) 09402100 (23) 1905-12/1922
   b) 09402000 (22) 1905-10/1925
   c) 09415000 (24) 1905-10/1909 **Note: incorrect data points found 10-12/1905
   d) 09421000 (25) 1905-12/1921
   e) 09422500 (26) 1905-12/1950 and 5/1951
   f) 09426000 (27) 1905-12/1913
   g) 09427500 (28) 1905-12/1934
   h) 09429490 (29) 1905-12/1934

   *This data was removed from further analysis. It will need to be estimated using appropriate techniques to reproduce a full dataset. Project staff should discuss the issue and select the techniques to be used, as well as which staff members will complete the extensions. **Note: See item 7 below regarding data extension.*

6) The single site in the lower basin that did not appear to have average monthly flows in the early part of the record is thought to have data generated using some other mechanism. This is because daily streamflow records appear to only extend backwards to 1923, according to the USGS website. Additionally, it is not a part of the historic calculations found on microfiche for the upper basin. The site is:
   a) 09382000 (21 – Paria River)

   Unless it can be verified that the data for this site was produced from gauged data, the record should be shortened and the data should be extended along with the sites above.

7) Data extension could prove to be difficult using conventional techniques. The data was analyzed for relationships that might be used for data extension through regression. Four sets of plots were produced to look at relationships between the sites in the study. These included plots of each site against the other basin sites for total natural flow, cumulative natural flow, localized inflow, and cumulative local inflow. The most important plots were the relationships between the localized inflows, however, since these would be the relationships of interest for producing appropriate regression models. The plots showed that the localized inflows from the sites with the short records do not relate well to each other, nor any other sites in the basin. Most of the linear relationships have an R-squared value of less than 0.2, with several sites having a near zero R-squared value. A review of a plot of the last 5 years of local flows at the sites with short records also shows that there does not appear to be a relationship between the sites. This strongly suggests that no other sites in the project will be able to be used for extension of the data.

   To compound the problem, there do not appear to be any other unregulated streamflow sites in the vicinity with long records. Without any real relationships

between the project sites or any other local sites, the use of regression models may prove difficult if not impossible. Additionally, plots were created to look at autoregressive relationships in the sites with short records. The plots showed that there isn't a strong autoregressive nature to any of the local inflow datasets. The R-squared values associated with lag-1 curves was generally less than 0.5, and a similar result was found with lag-12 (1 year) curves.

*The project staff needs to discuss this issue. Due to the difficulty in finding any real relationships between the local inflows in the basin, there could be a major problem extending the records using mathematical models. One option that comes to mind is modeling the total natural flows, however it is questionable whether this would actually produce results that would improve the dataset. The alternative is the actual re-computation of historic natural flows, however it is questionable that the data exist to attempt such a project, given that it wasn't done before. Additionally, staff time would likely become an issue with this method. **Note that the lack of relationships between data from various sites can be reviewed in the "CONaturalFlows_RelationshipPlots.xls" spreadsheet.*

3.3 – Negative and Zero Values

8) Negative total natural flows were found in several sites. In talking with Jim Prairie, the reason that negative total natural flows were left in the existing recordset is because the RiverWare models do not have a problem modeling with negatives, and the negative values were attributed to uncertainty in the calculations. Negative total natural flows were found in following sites after 1951 (years with apparent non-averaged data)
   a) 09302000 (14) *(headwater)
   b) 09328500 (17) *(headwater)
   c) 09355500 (18) *(headwater)
   d) 09379500 (19)
   e) 09421000 (25)
   f) 09422500 (26)
   g) 09427500 (28)
   h) 09429490 (29)

The negative total natural flows may not be desirable in the dataset that will be stochastically modeled, thus these negatives should be removed. The following is a summary of the suggested changes to rectify these negatives:

Sites 25 – 29

There is a single month of negative flows in this time period (4/67), which is due to a large intervening negative in site 25. Visual inspection indicates that the problem appears to be a bad data point in the record for site 25. The logical solution would be to inspect the historic record if possible, or estimate the intervening flow using interpolation or regression to other sites.

Sites 14-19

These sites have the following number of negative flows in the post-1951 dataset:

14: 2 points (9/83, 8/85)

17: 2 points (10/67, 10/77)

18: 6 points (8&9/78, 9&10/79, 7&8/94)

19: 3 points (8/78, 1/81, 11/93)

Each of the sites except 19 is a headwater node. The following are the suggested causes and solutions for the negatives:

14: The 9/83 event appears to be a large spike followed by a large dip, including the negative. This corresponds to a known period of extreme runoff. The suggested solution is to review historic record and fix if possible, or even out the extremes by raising the negative to zero and reducing the spike by a corresponding amount. The 8/85 event appears to be anomaly, and may just need to be adjusted to zero.

17: Both events appear to be slight dip below normal. Unless historic data can be used to improve number, the suggested solution is to raise the value to zero.

18: All three pairs of events follow extreme flow spikes. This could be an artifact of calculations such as for bank storage. The negatives are not extreme, however, thus the suggested solution is to raise the negatives to zero or slightly higher, and reduce the peaks by an equivalent amount.

19: The 8/78 event appears to be caused by corresponding event in site 18. This should automatically be adjusted when site 18 is fixed. The 1/81 event is caused by dip in the corresponding intervening flow. The suggested solution is to raise the intervening flow by appropriate amount. The 11/93 event is caused by a large dip in the corresponding intervening flow. The suggested solution is to raise the intervening flow, and potentially decrease the slight spike preceding it by an equivalent amount.

9) Zero total natural flows were found in following sites after 1951
   a) 09402000 (22) *(headwater)
   b) 09426000 (27) *(headwater)

   Due to the fact that these are headwater nodes in an arid region, it is not unexpected to find zero-valued data. The zeros are not a concern, however they both are sites where averaged data was encountered and subsequently removed, thus the effect of the zeros will need to be considered when selecting a method to extend the recordsets.

10) Many sites have a large number of negative and zero-valued intervening flows. The ones with the largest number (>100) of negative periods include:
   a) 09180500 (8)
   b) 09217000 (10)
   c) 09315000 (16)
   d) 09380000 (20)
   e) 09402100 (23)
   f) 09421000 (25)
   g) 09422500 (26)

h) 09427500 (28)
i) 09429490 (29)

These values are not a major concern. Many reaches in the Colorado River system may legitimately be losing reaches, thus with the exception of the periods where these cause the total natural flow in the river to drop below zero, they should be used as-is.

The ones with the largest number (>100) of zero-valued periods include:
j) 09402000 (22) *Headwater

These values are not a major concern. Since this is a headwater, zero-valued periods can be expected.

3.4 – Extension Beyond 1995

11) The extension of records to WY2000 is expected within the next couple of months. It is expected that they will have similar issues to the 1971-1995 records (negatives, zeros, etc), since they are being calculated using the same methods.

No action is necessary. The updates will need to be obtained and analyzed as soon as they become available.

12) Extension beyond 2000 would be difficult because CU&L (consumptive use and loss) reports only come out every 5 years (these are the reports upon which a portion of the upper basin natural flow calculations are based). The demands would need to be extrapolated forward from 2000 (assuming that report is available… 1995 may be the last available report at the moment). To extrapolate forward, the CU&L would need to be estimated for each of the 21 upper basin sites, and it is likely that each of the 8 components of the CU&L calculation would need to be considered separately (irrigation, reservoir evap, stockponds, livestock, thermal power, minerals, M&I, export/import). Additionally, decree accounting records would need to be obtained for lower basin users.

It is suggested that no effort be made to extend the data beyond 2000. With a limited amount of staff time available, the resulting data would likely have an extremely large measure of uncertainty associated with it, particularly due to the extreme nature of the climate in the basin over the last 4 years.

3.5 – Nonstationarity

13) Longer term nonstationarity may exist in the basin. Proxy data throughout the West is starting to show that there may be long term trends in water supply (an example is shown in Figure 1), and some trends may exist in the natural flow calculations.

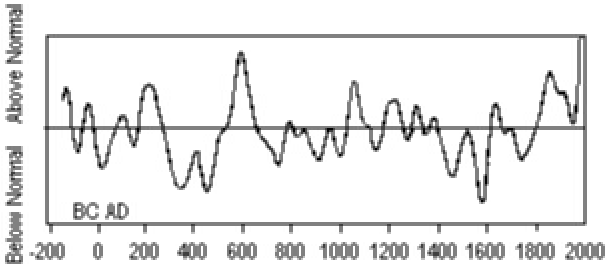


Figure 1. - Reconstruction of annual rainfall from El Malpais National Monument, New Mexico, based on tree-ring analysis showing long-term (more than 100 years) climatic trends. (Adapted from Grissino-Mayer 1997, taken from http://www.cpluhna.nau.edu/Change/climate.htm).

In an effort to identify trends in the dataset, four sets of plots were generated. These included total flow hydrographs, localized inflow hydrographs, and mass curves (cumulative flow) for both total flow and local inflow. Rapid visual analysis revealed the following results;

Trends in total natural flow

The 10-year moving average curve of total natural flow in the upper basin shows slightly higher flows in the early records (through early 30's), then lower flows from the 30's through the early 80's, followed by a wet period in the mid to late 80's. The shorter lower basin records show no distinct trends other than the same wet period in the 80's.

Trends in local inflow

The upper basin local inflow calculations show similar trends to those observed in the total natural flow calculations. One exception is node 8 (09180500), which illustrates wavelike trends from approximately 1966 through 1995. The shortened lower basin local inflow data show a variety of different apparent trends, most notably on nodes 26 and 29. Both of these nodes have a visible wavelike trend, however the trends do not seem to correspond to each other or any other noticeable trends in the basin.

Trends in cumulative total natural flow

Changes in slope of cumulative flow generally indicate a changing trend in the data. Visual analysis of the cumulative total natural flow charts indicates that in the upper basin, most sites seem to have a change in curve slope in the early 30's, and again near the late 80's. These trends correspond to those seen in the total natural flow hydrographs. Due to the flashy nature of some of the lower basin sites, it is difficult to visually determine if there is a change in slope for these shortened datasets. Some of the mainstream sites appear to have a change in the late 80's, however this change is not as well-defined as the changes seen in the upper basin.

Trends in cumulative local inflow

The changes in slope for cumulative local inflow in the upper basin that are visually noticeable correspond to those seen in cumulative total natural flow. As with the hydrograph for local inflow, the trends seen for site 8 are drastically accentuated, with a major period of negative local inflows in the 70's. As with the trends seen in the local inflow hydrographs for the

lower basin, the cumulative charts show drastically varying trends, ranging from an S-type curve on site 26 to a wavelike curve on site 29.

The trends should be reviewed by project staff by looking through the charts located in the Appendix. These trends should be discussed, and a decision should be made as to whether they should be addressed or not. The primary reason that they might need to be ignored is that any trend removal might be difficult to justify, given that actual trends may exist in the basin, and these trends may not be artifacts of calculation methods. Additionally, it would be difficult and time consuming to review the original records to identify where the trends originated if they do actually come from the calculation methods.

4.0 - Suggested Enhancements to Existing Dataset

The following is an outline of the detailed suggestions above:

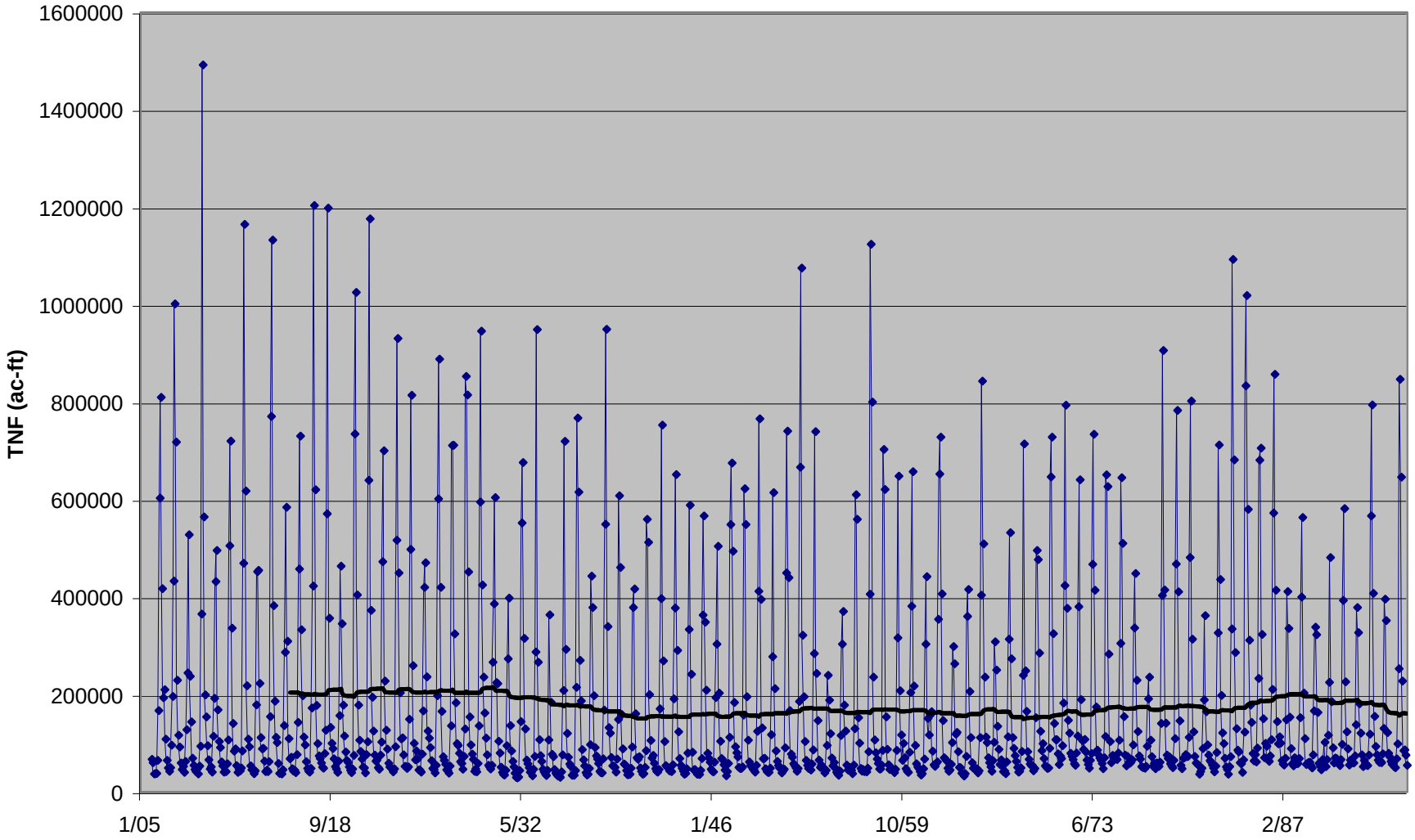
1) An enhanced map showing the subbasins should be generated to help with general project work
4) More time should be spent checking the microfiche records against the 'official' records, and a decision needs to be made whether to use the microfiche records or the 'official' records for San Rafael
5-7) The shortened records in the lower basin should be extended if possible, however a discussion should be held as to how this should take place. Currently there does not appear to be a good option for extension due to lack of relationships between the data for each of the lower basin sites.
8) The negative flows in the total natural flow should be removed using the procedures detailed in the section above
12) The data should not be extended beyond WY2000 due to the work involved in obtaining/generating the necessary data
13) Trends in the data should be discussed, however not removed due to the potential difficulty in identifying the source of the trends

5.0 - Conclusions

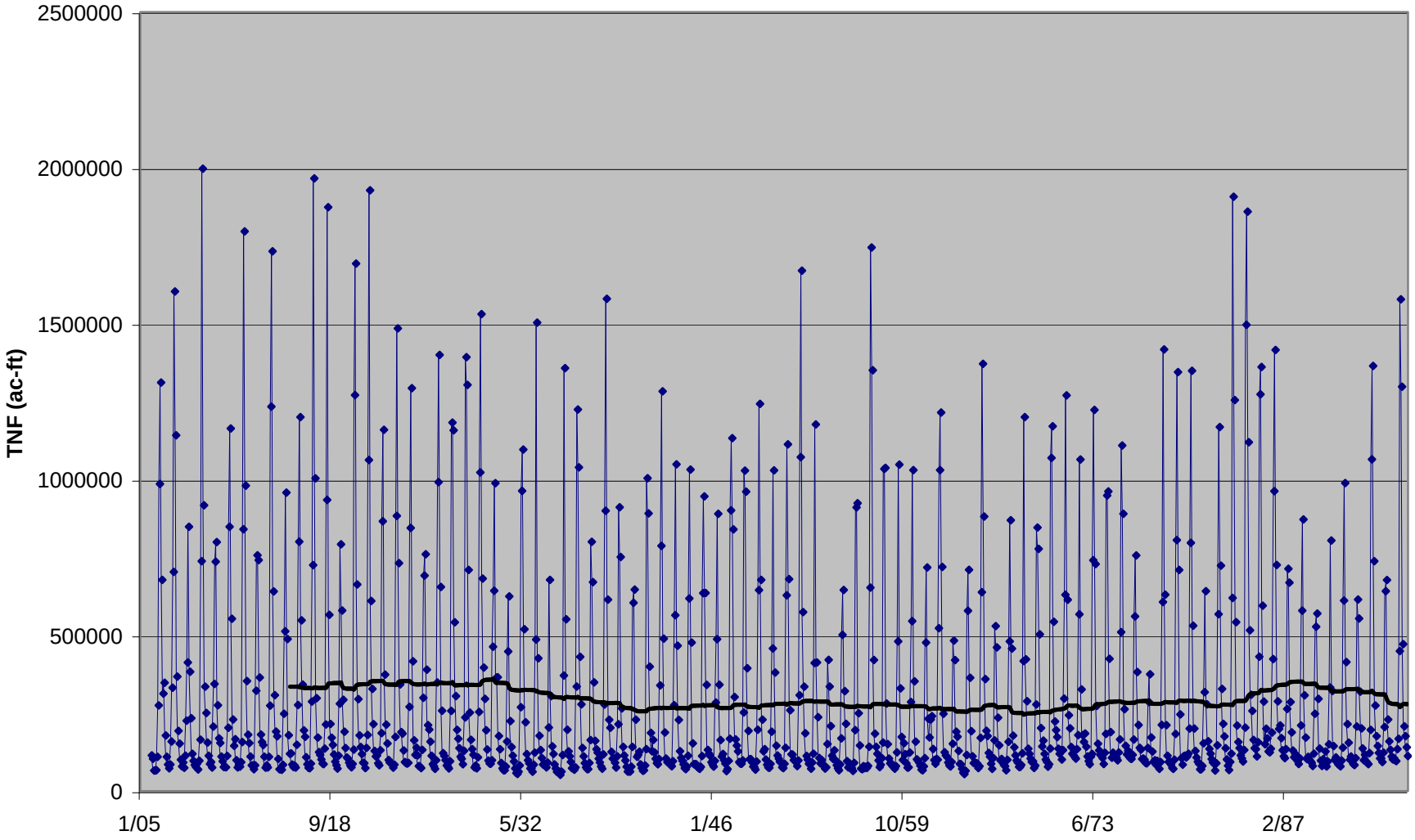
The project staff should review the above recommendations, and hold any necessary discussions. The changes should then be made to the dataset to obtain an appropriate base for the stochastic modeling portion of the project.

Total Natural Flow Hydrographs

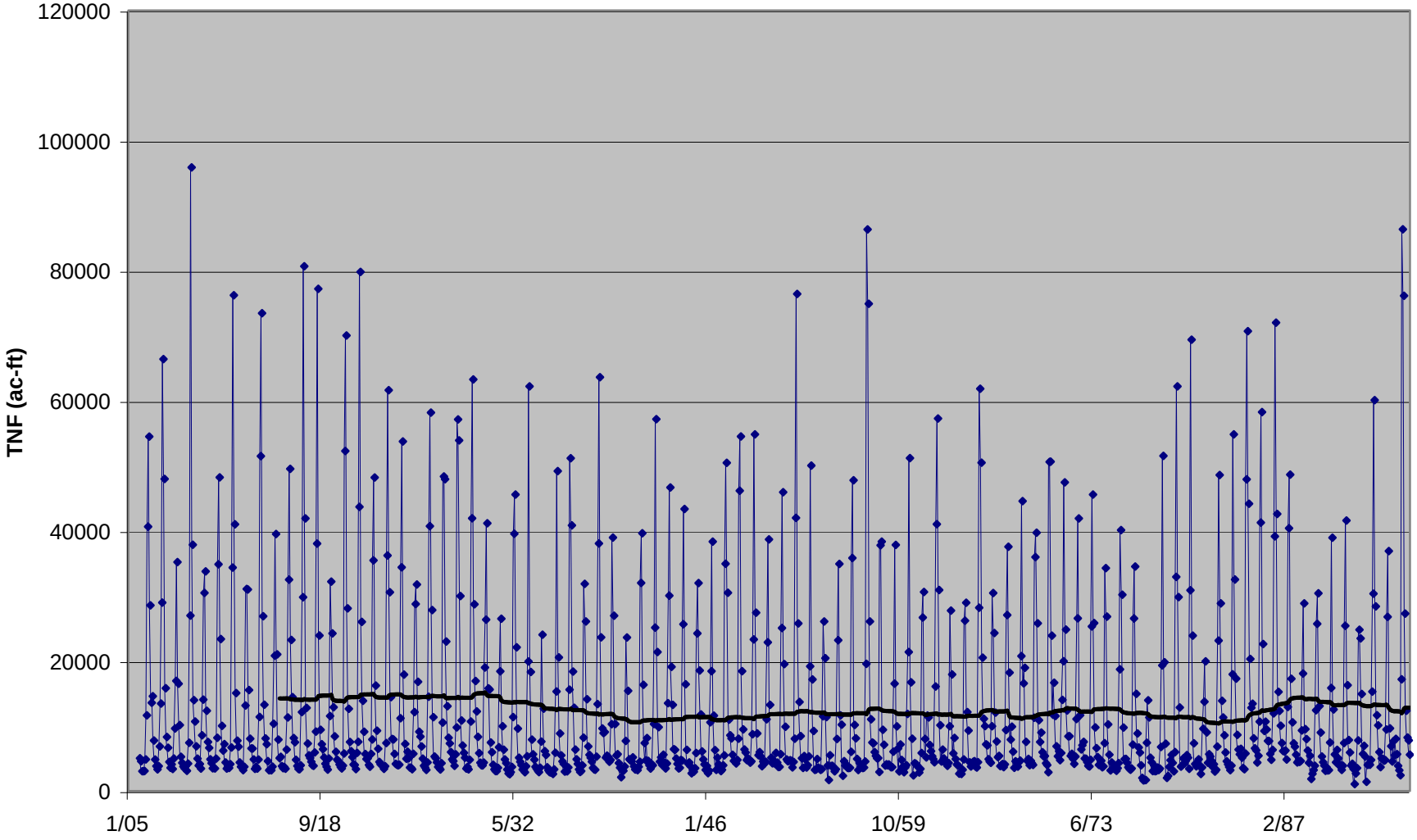
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 1



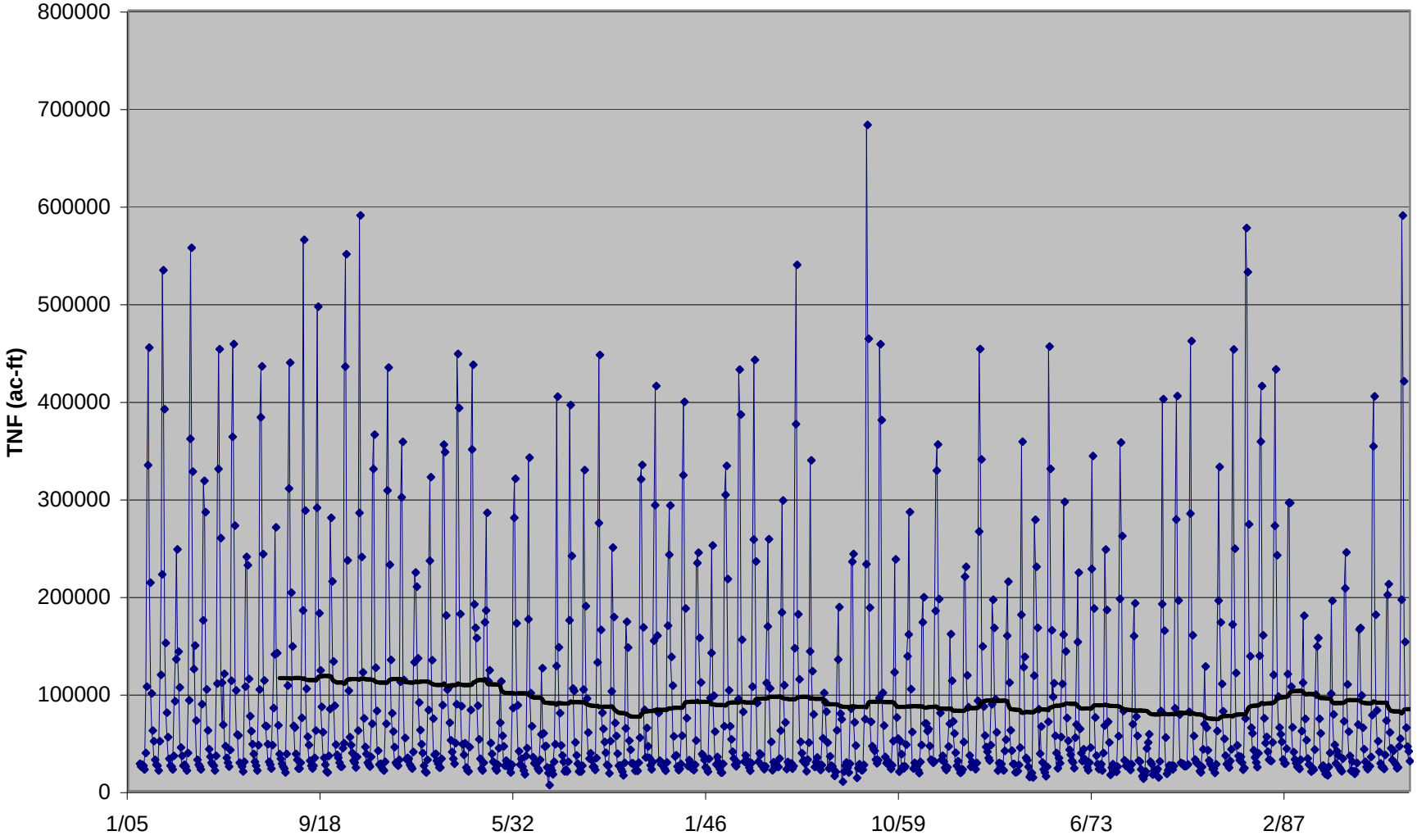
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 2



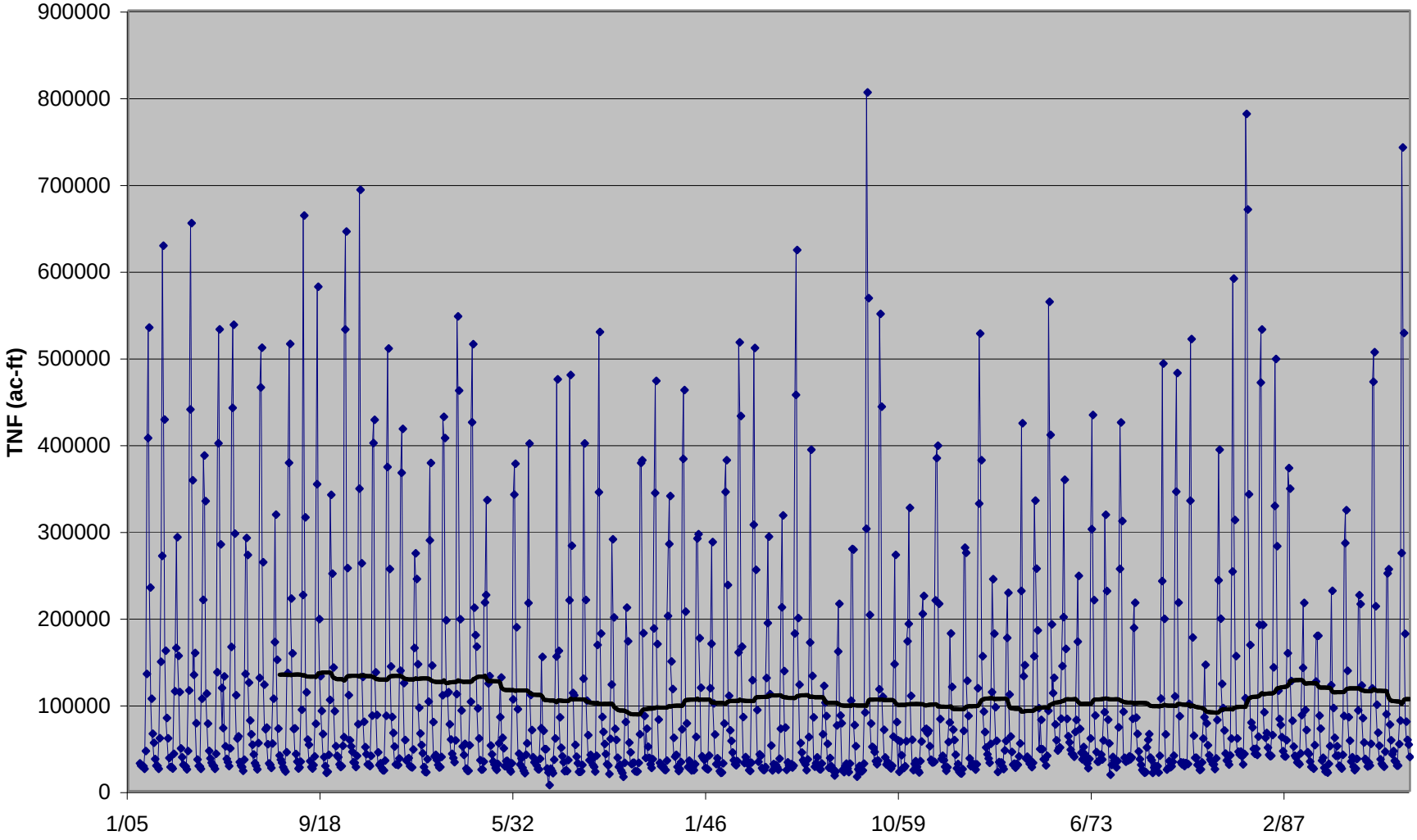
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 3



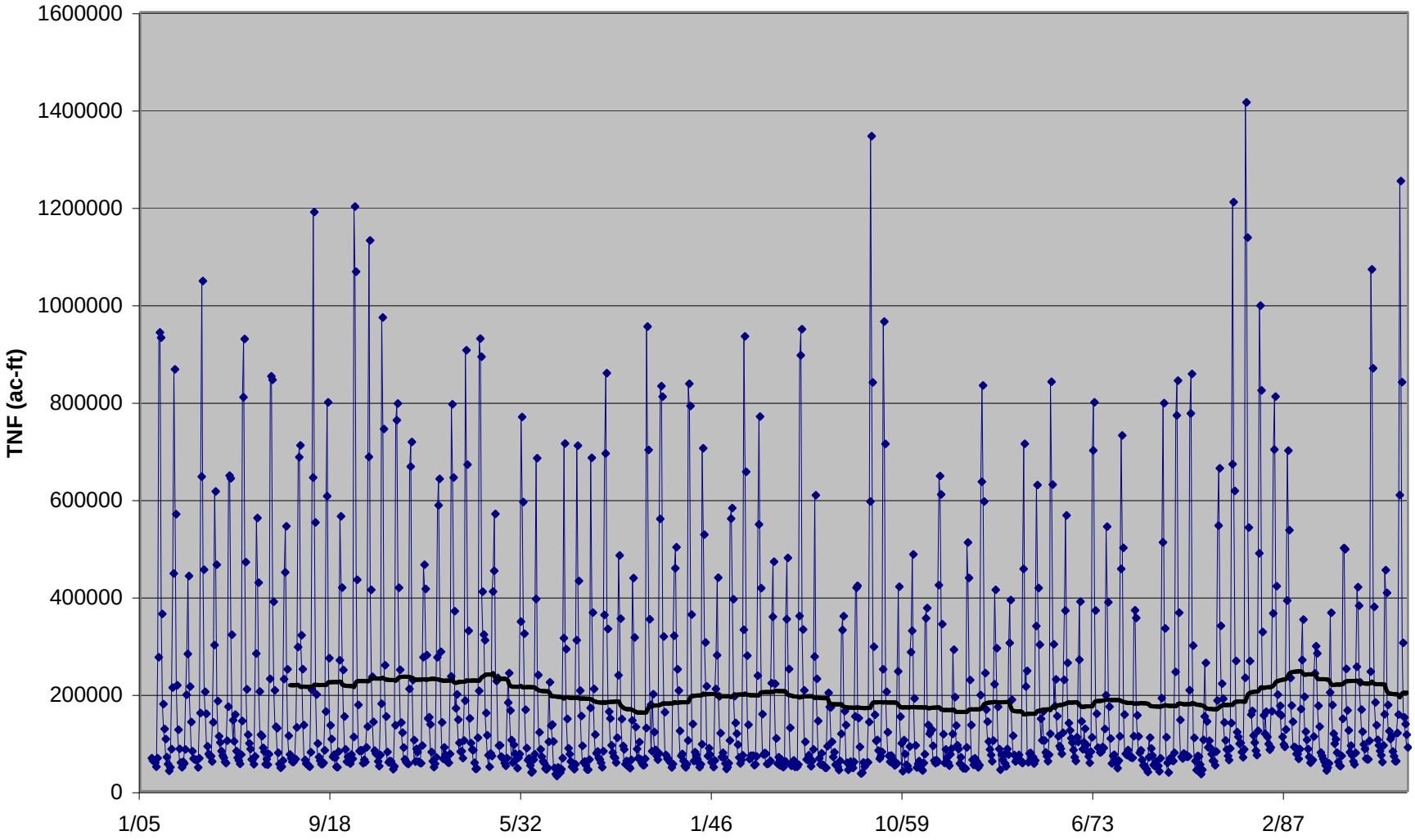
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 4



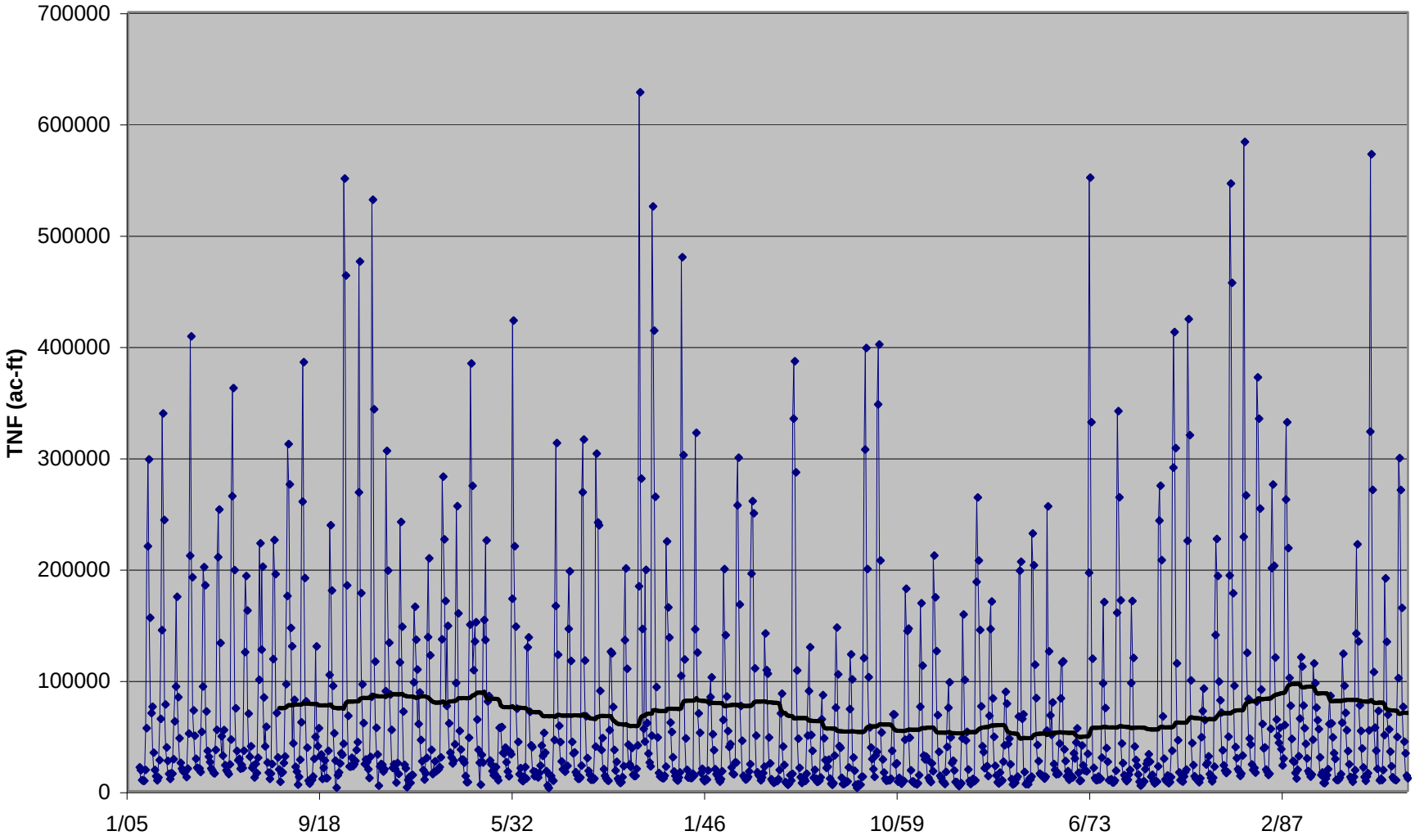
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 5



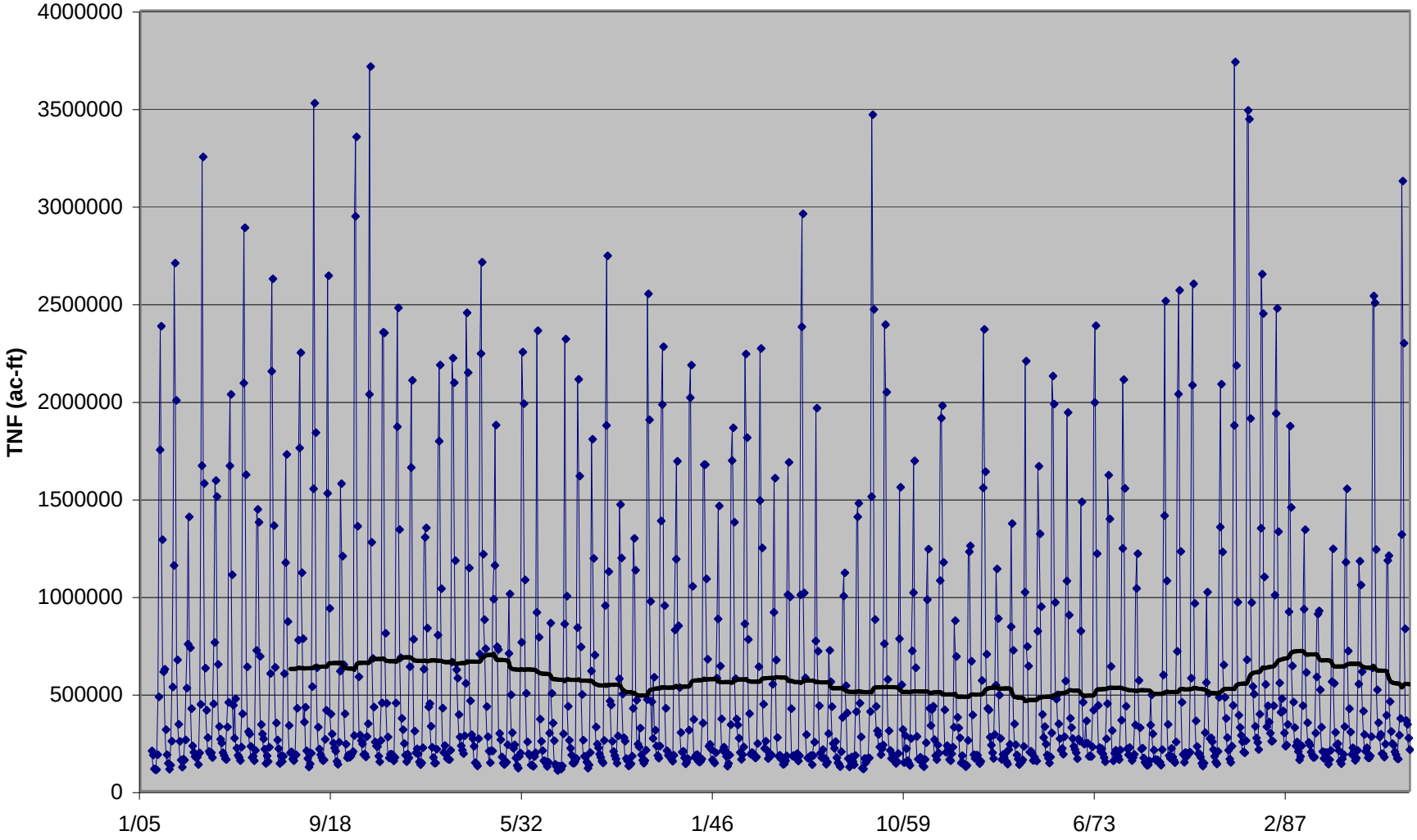
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 6



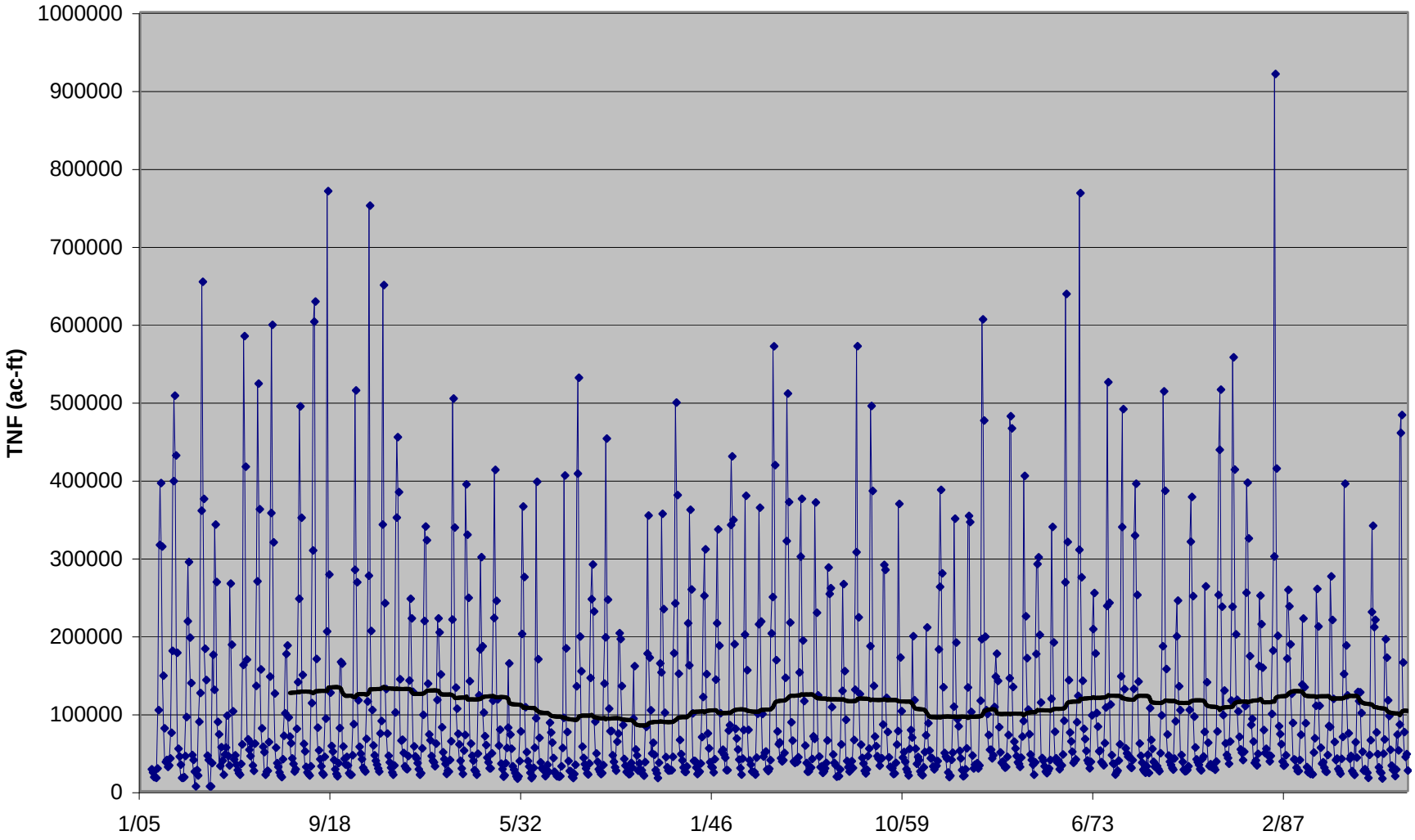
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 7



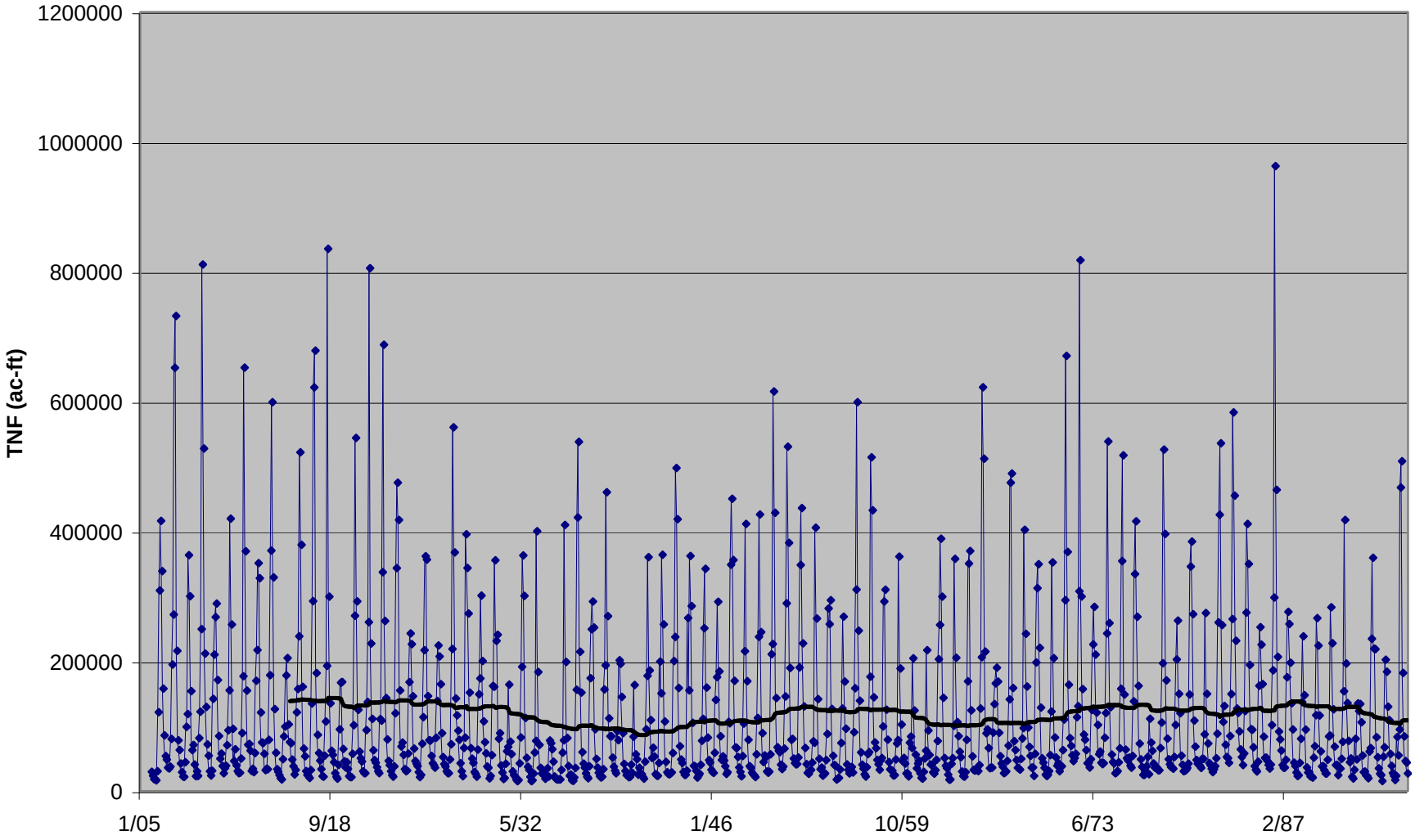
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 8



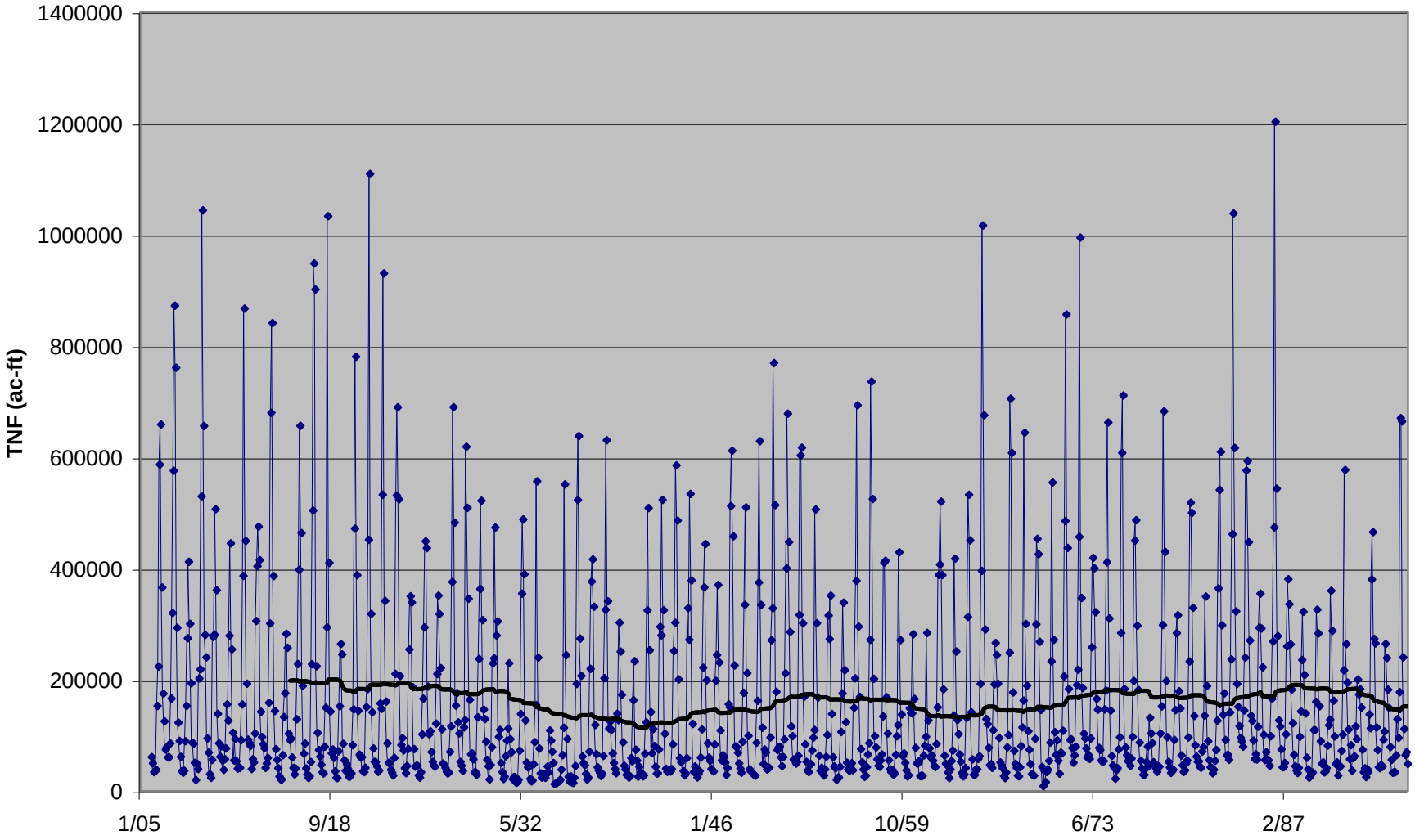
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 9



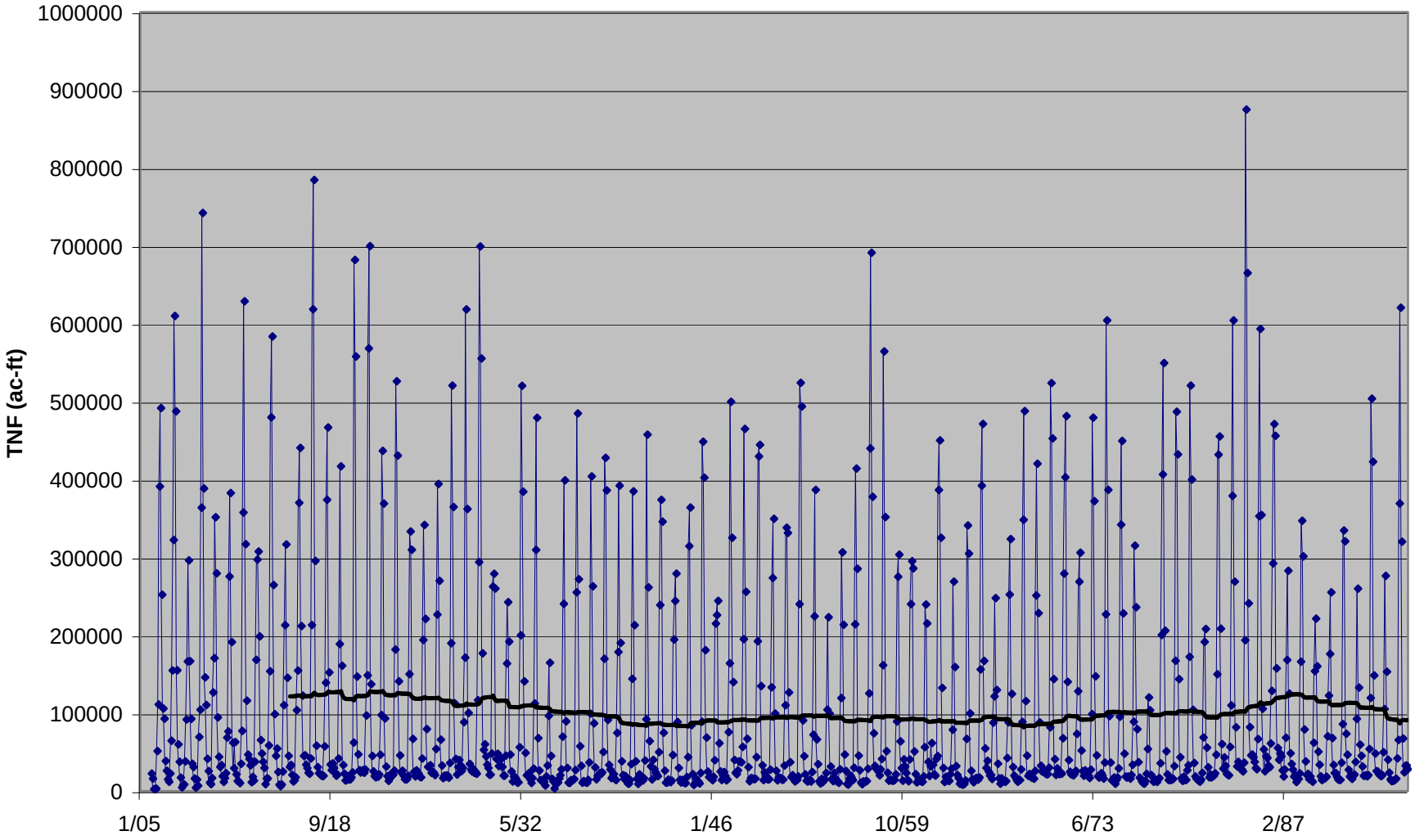
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 10



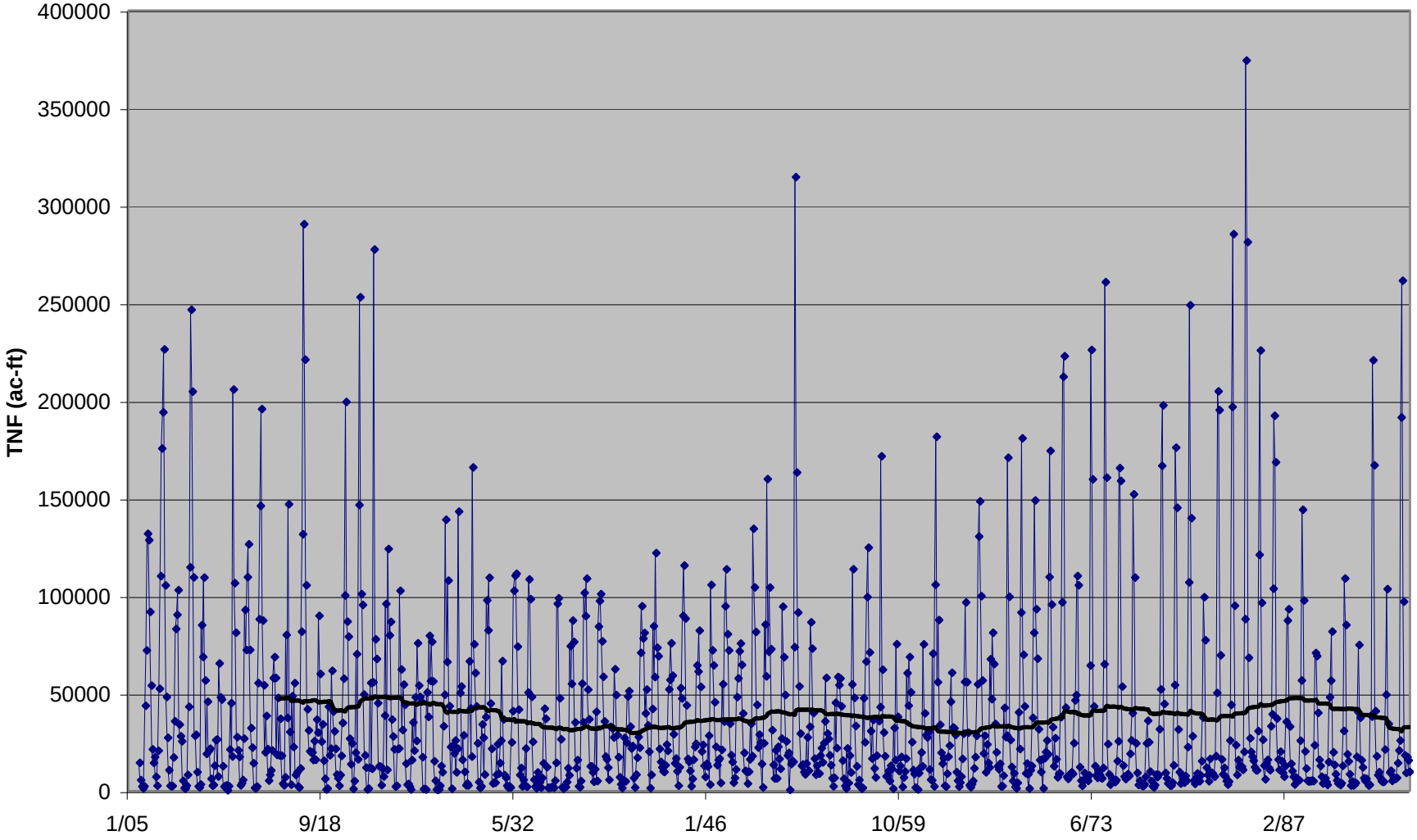
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 11



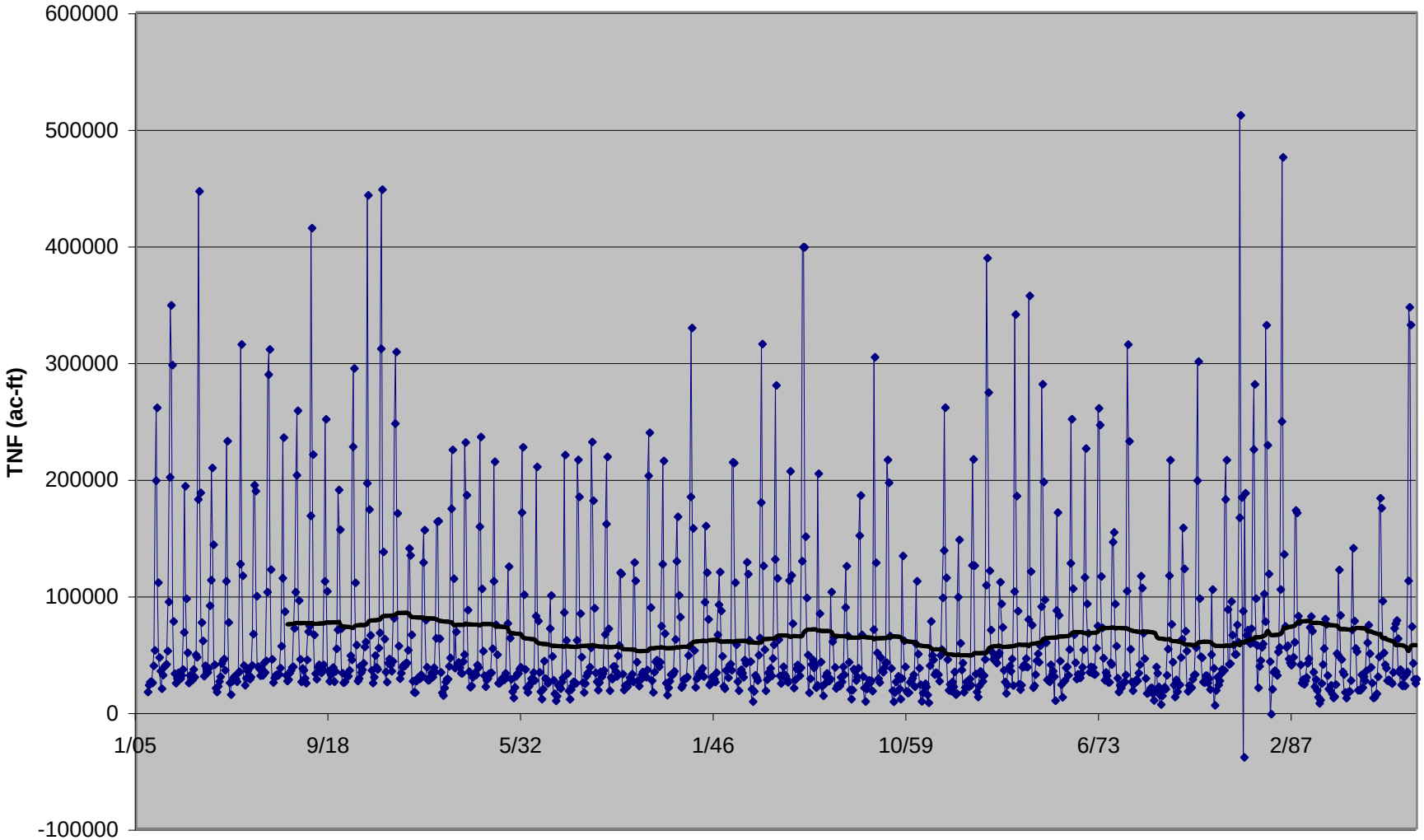
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 12



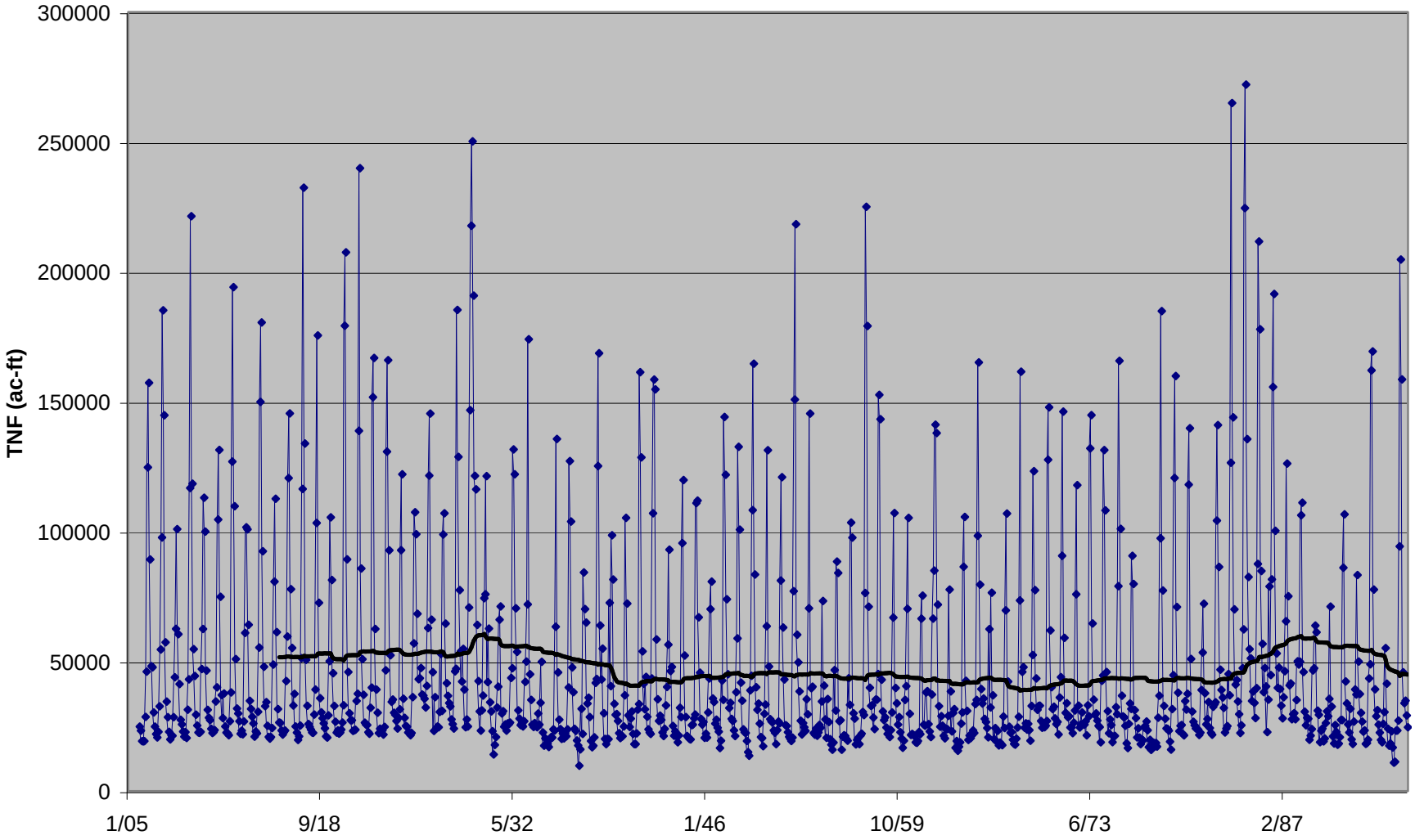
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 13



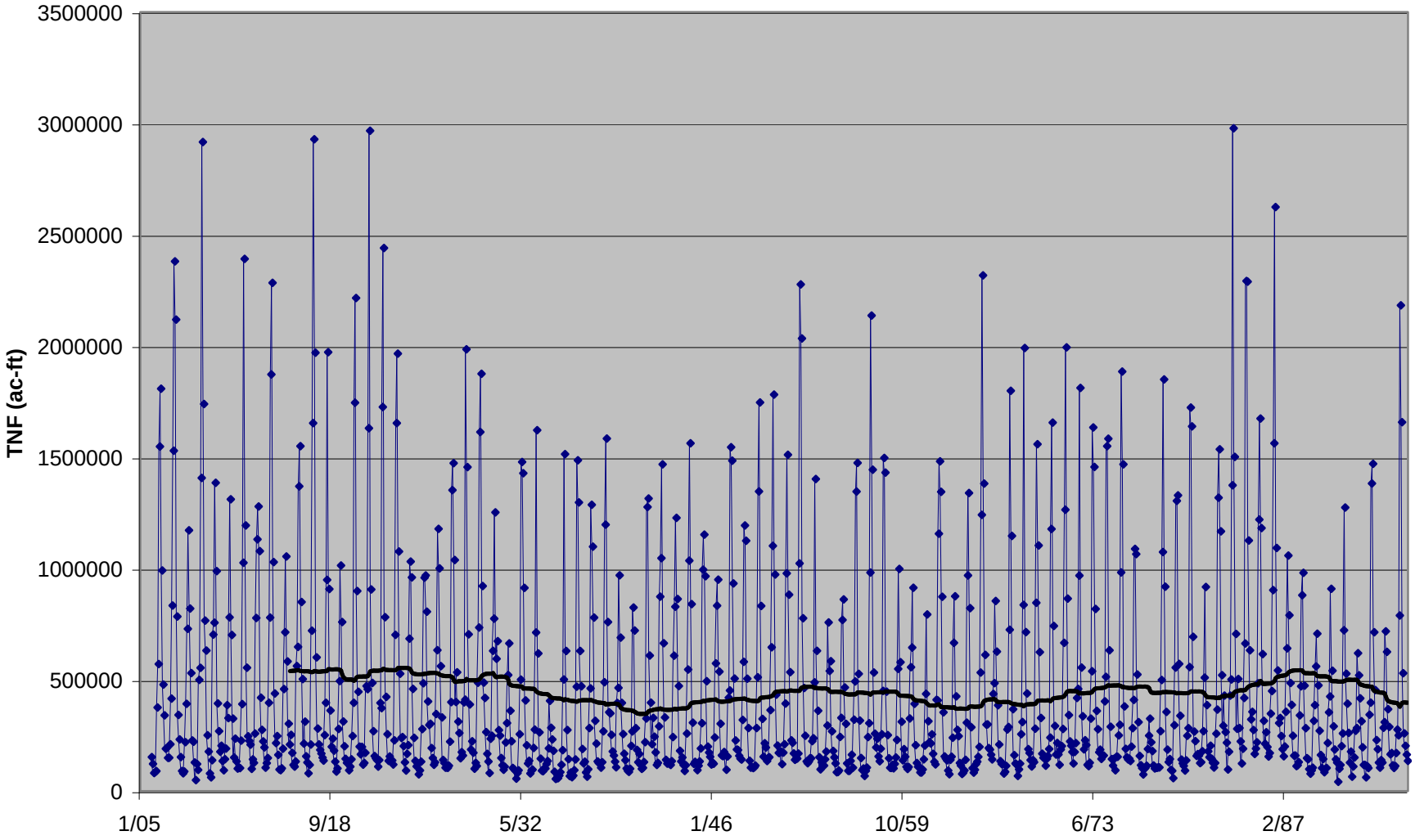
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 14



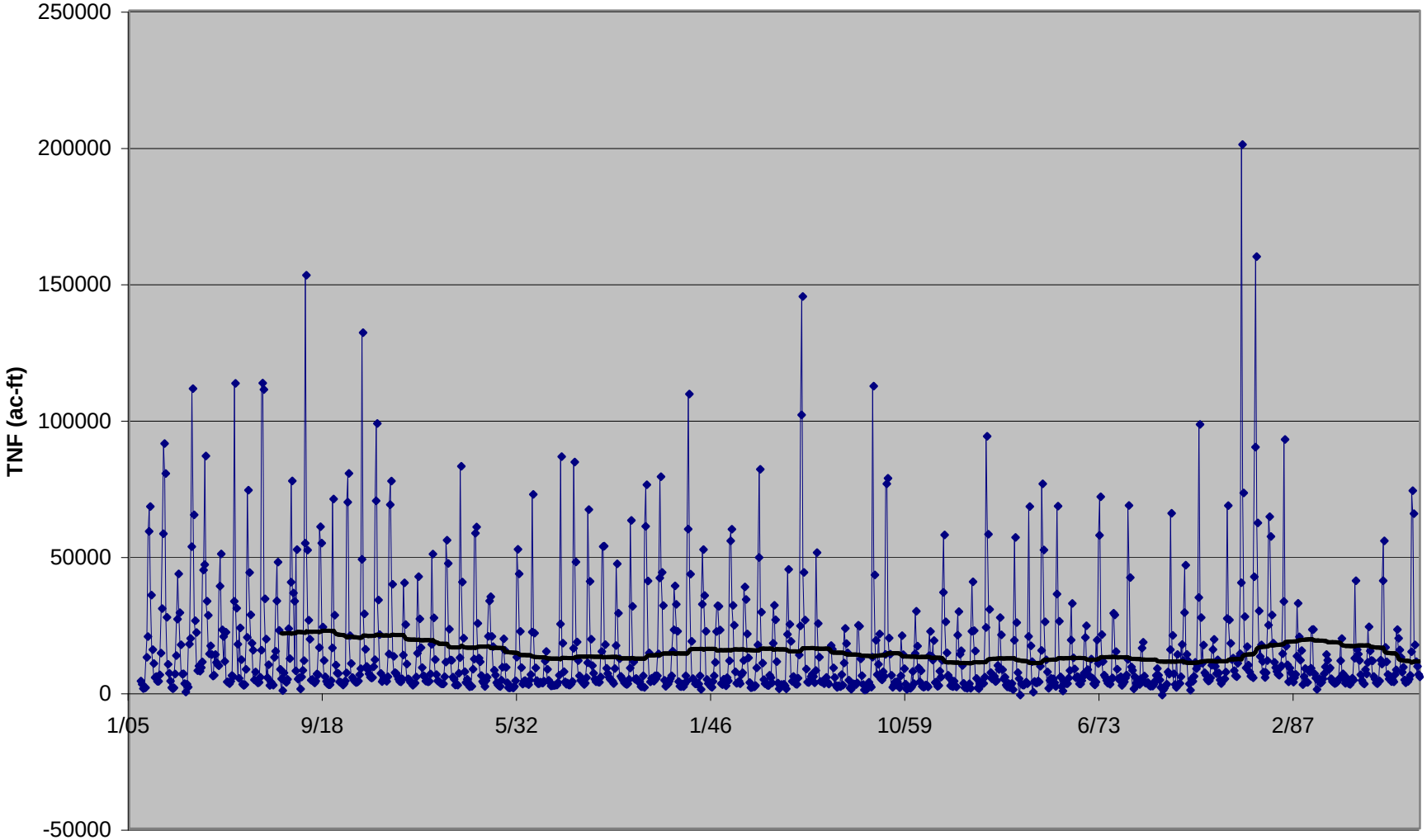
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 15



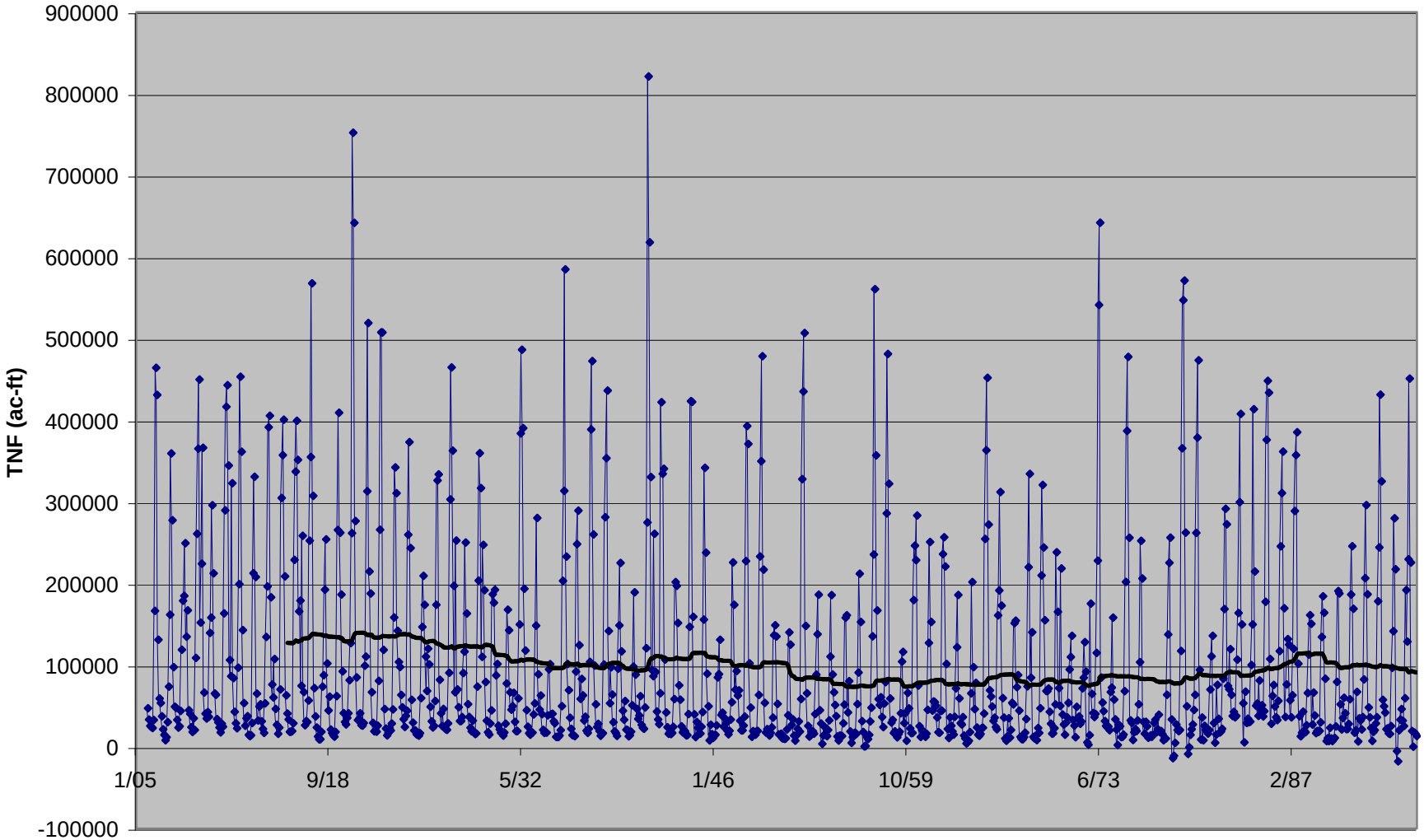
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 16



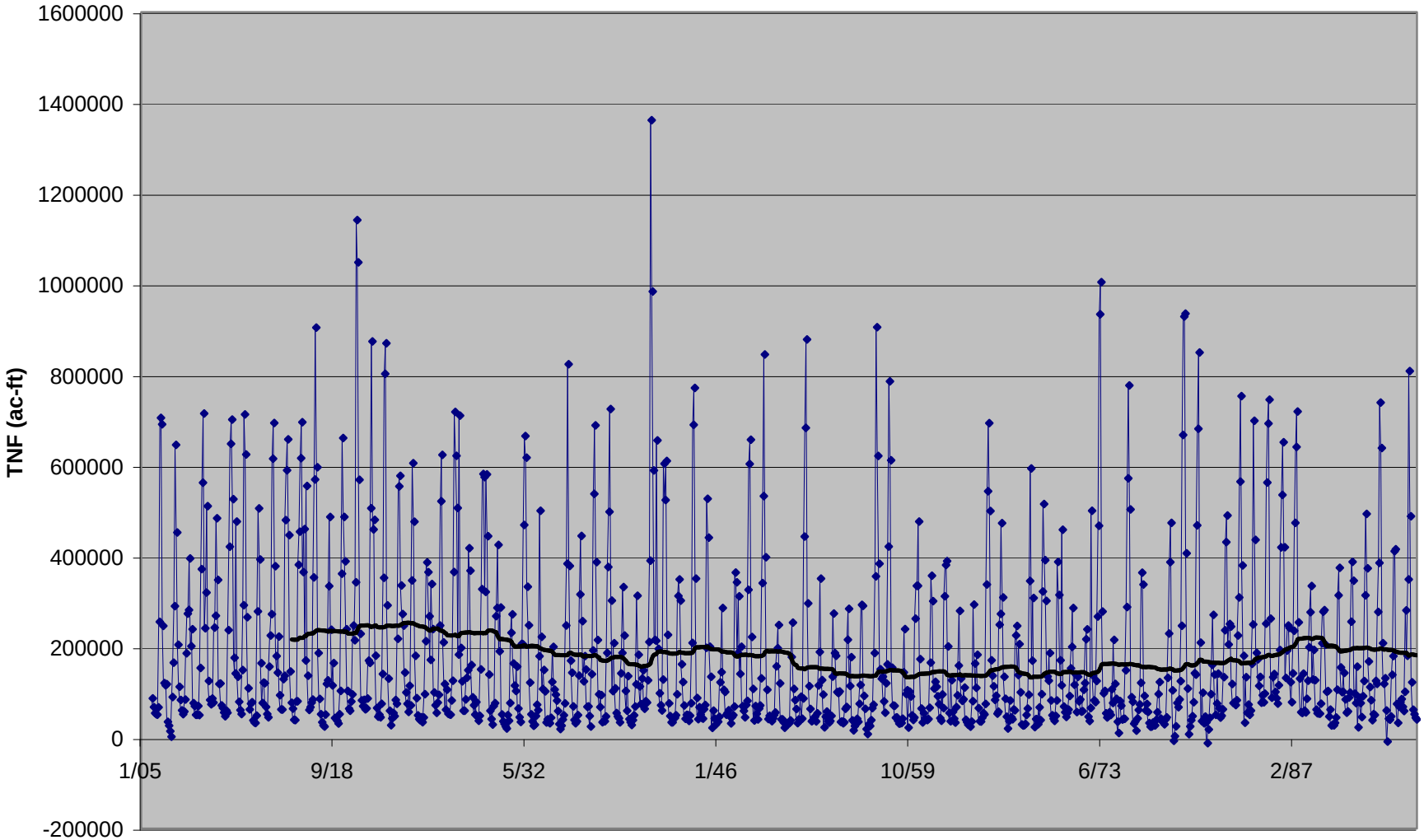
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 17



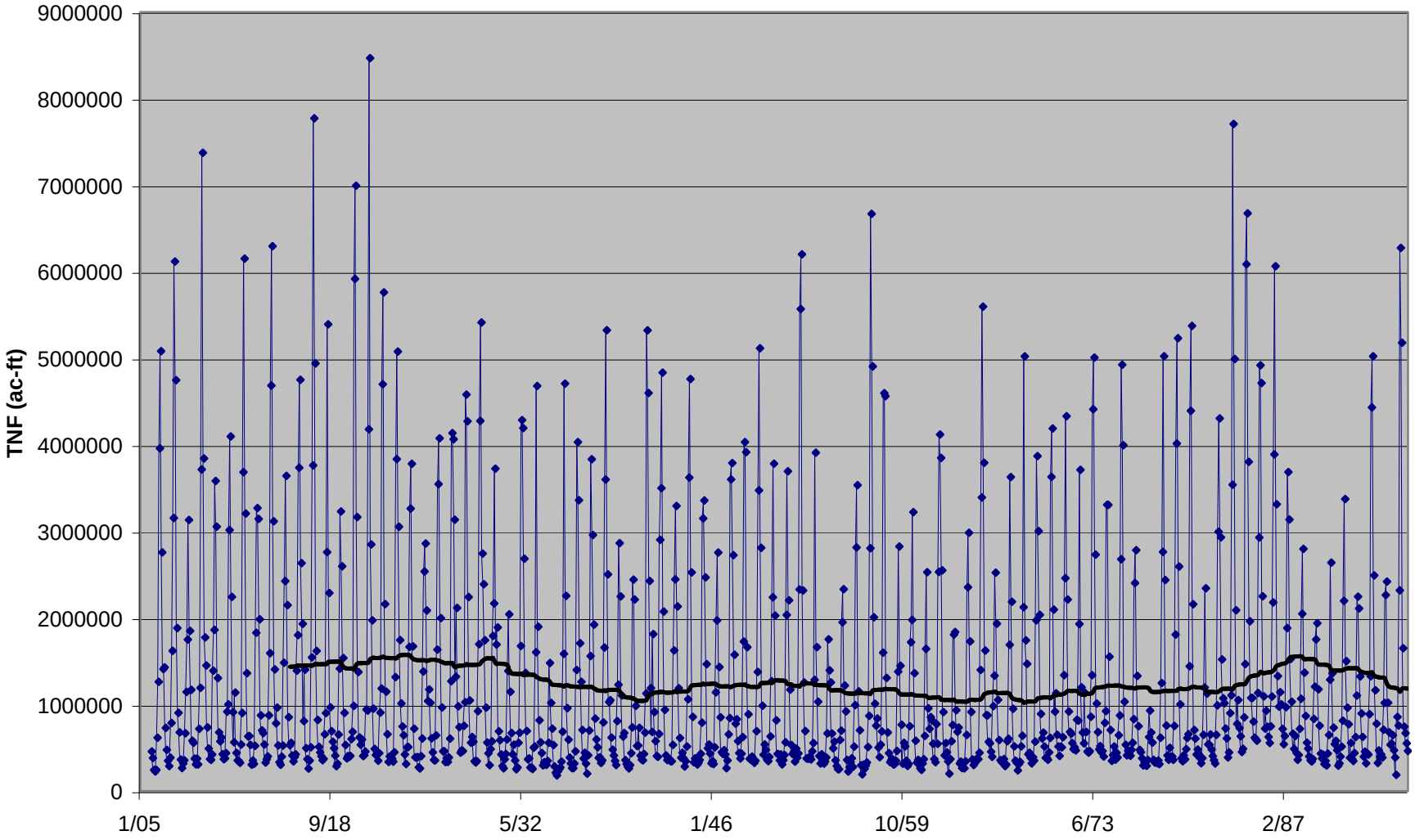
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 18



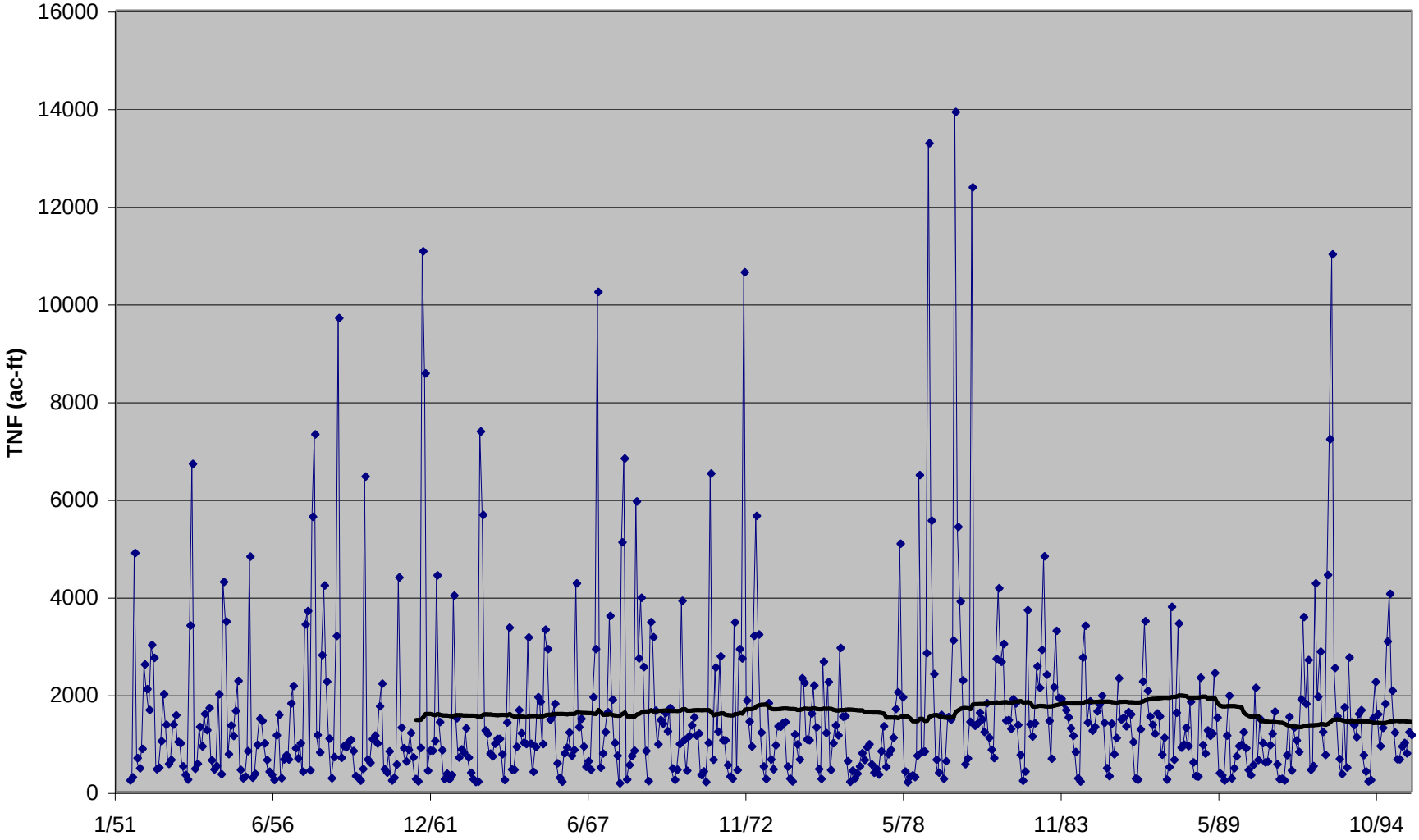
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 19



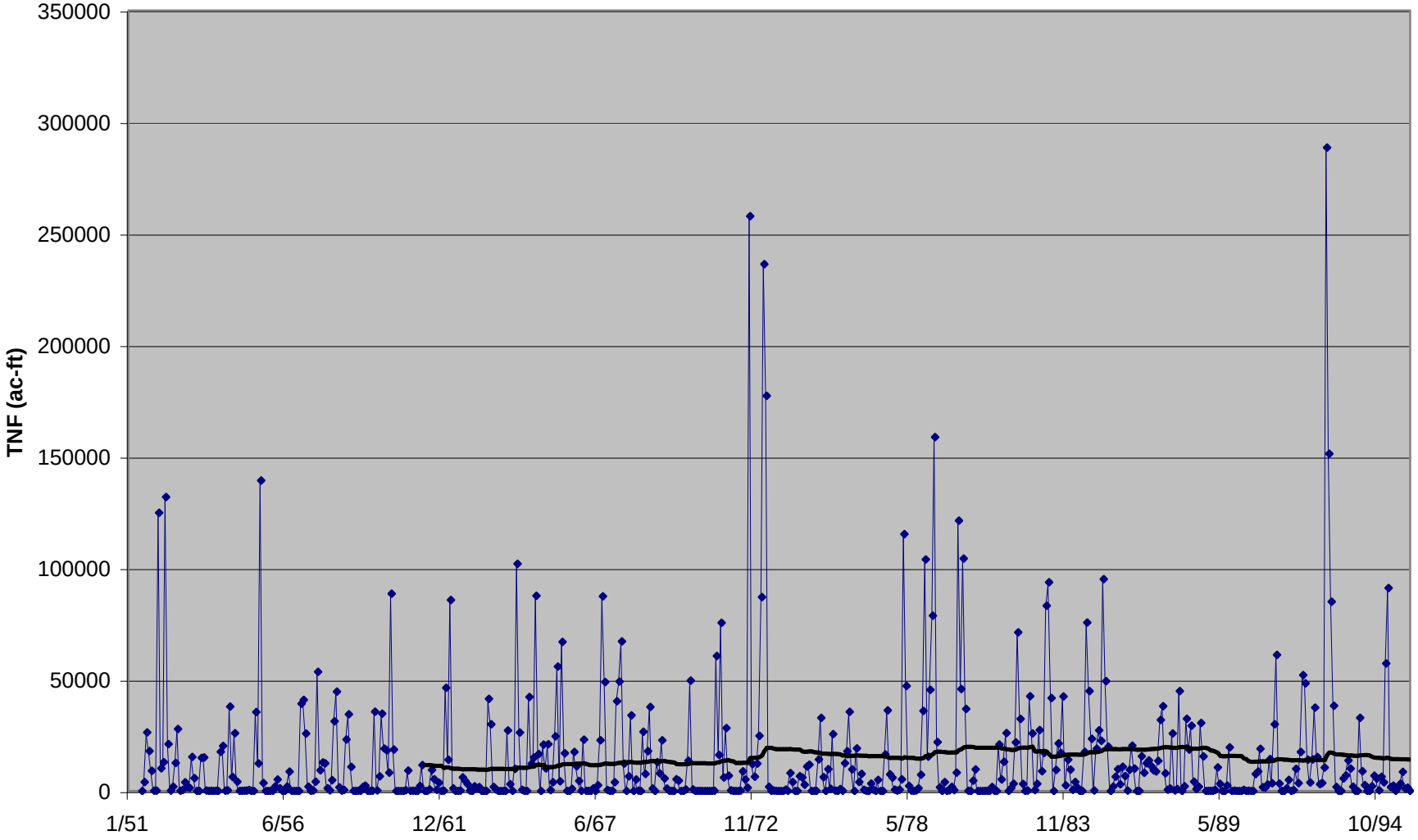
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 20



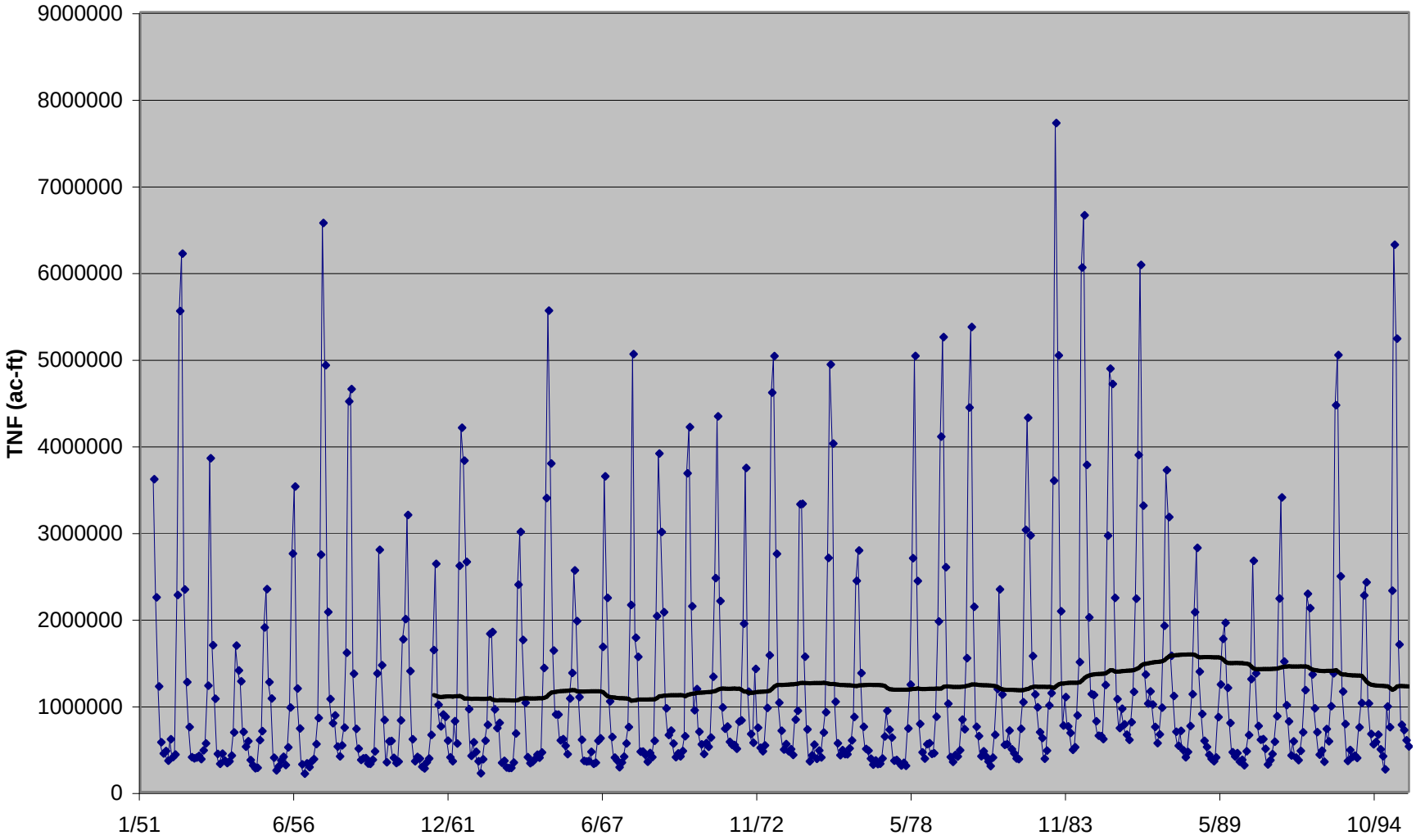
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 21



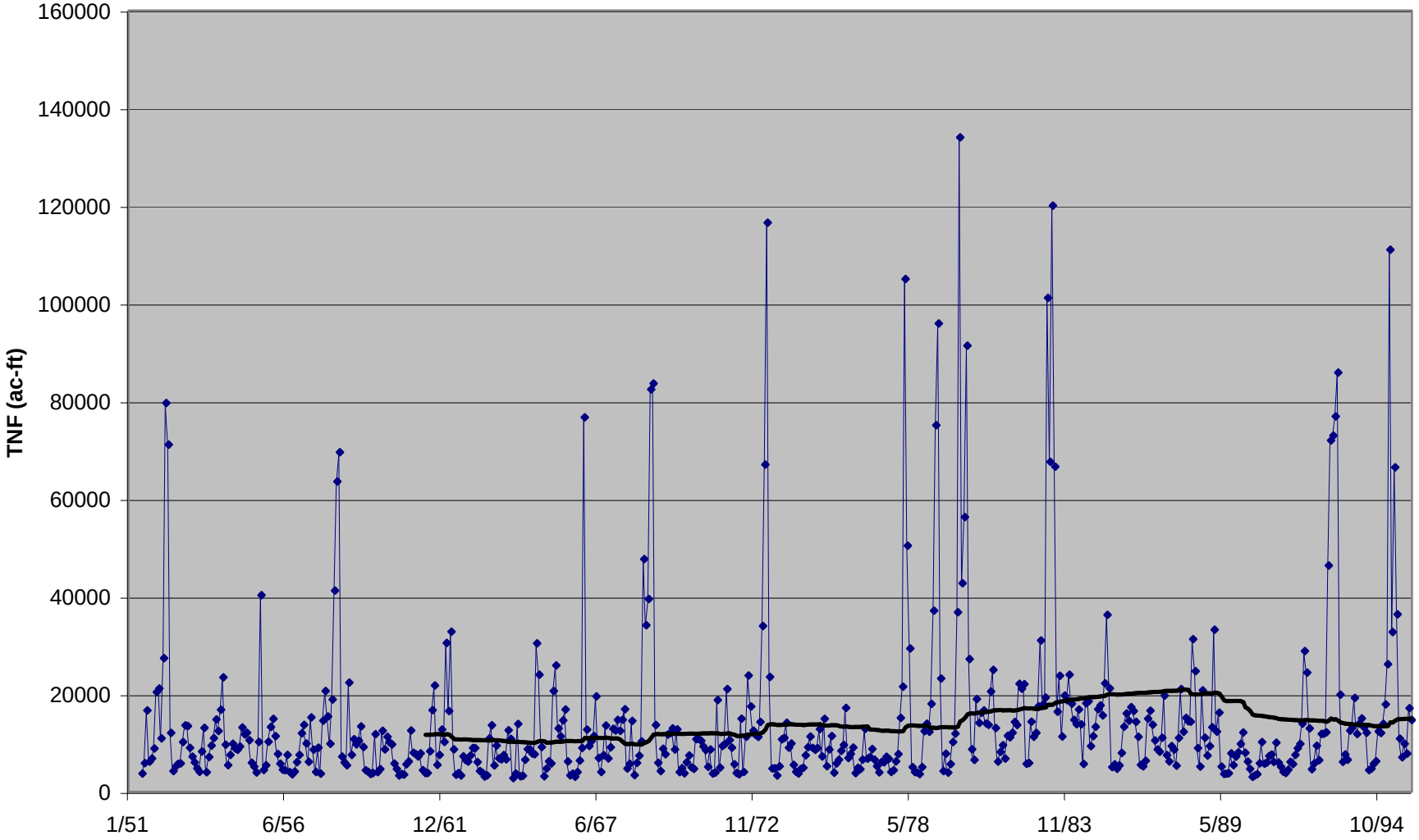
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 22



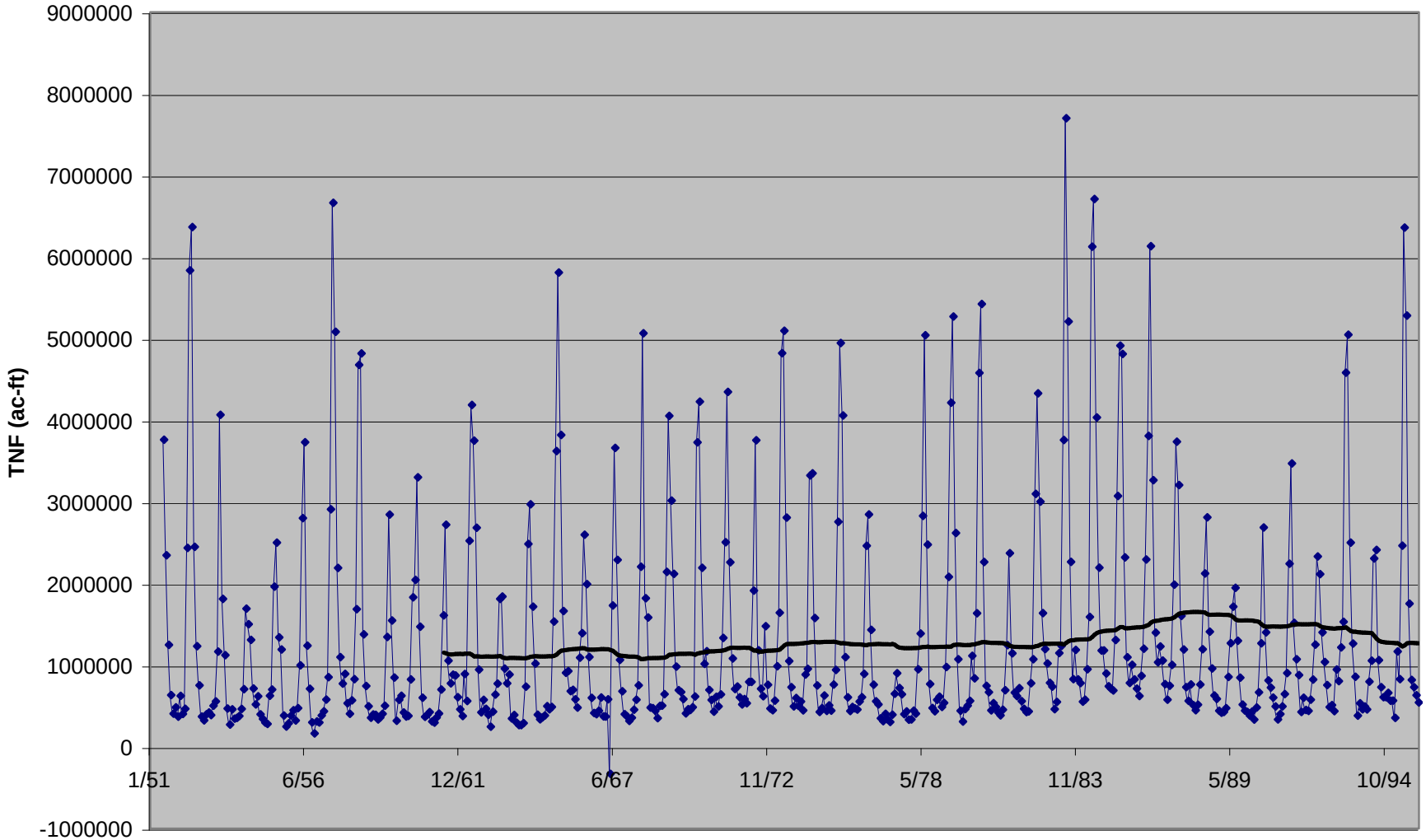
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 23



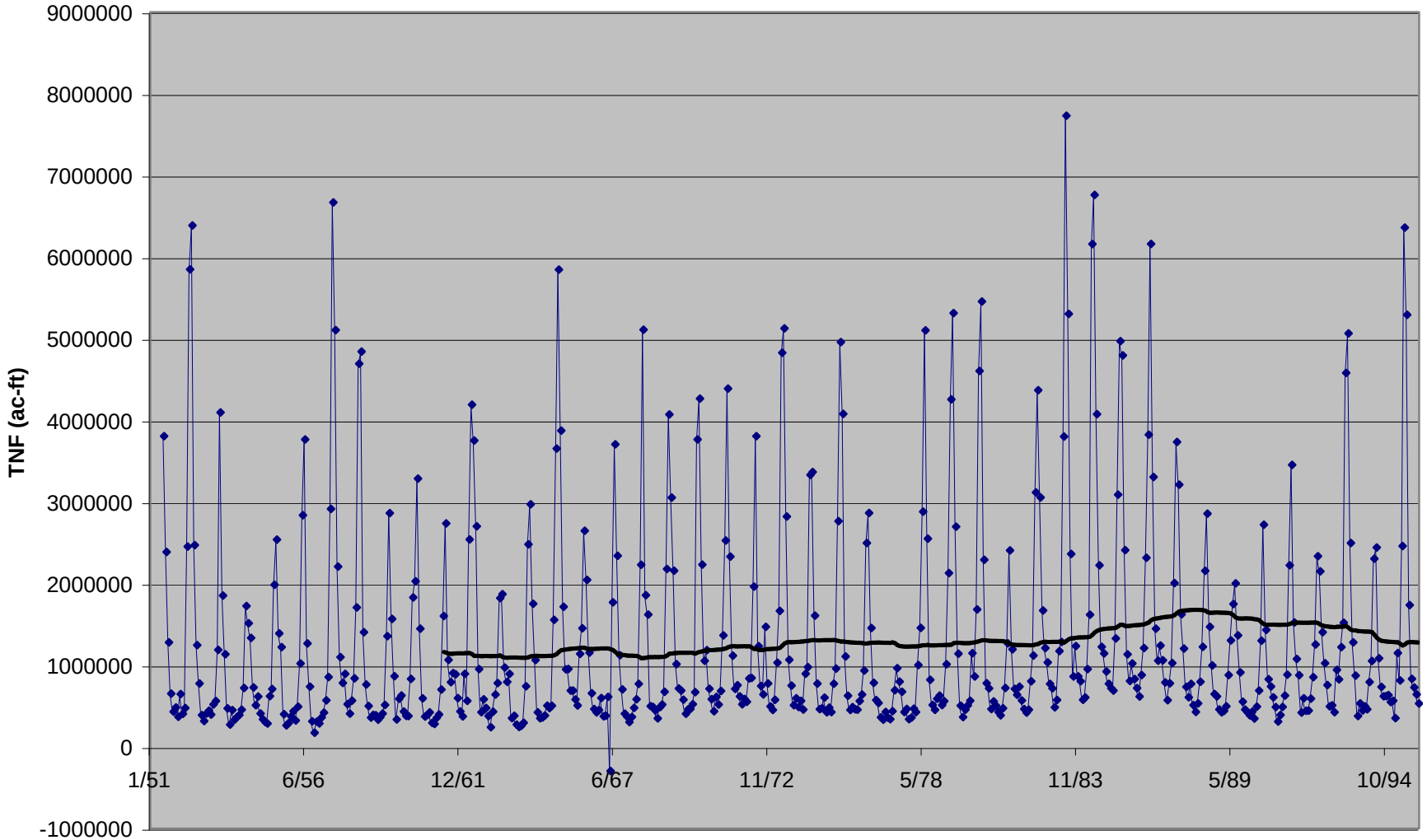
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 24



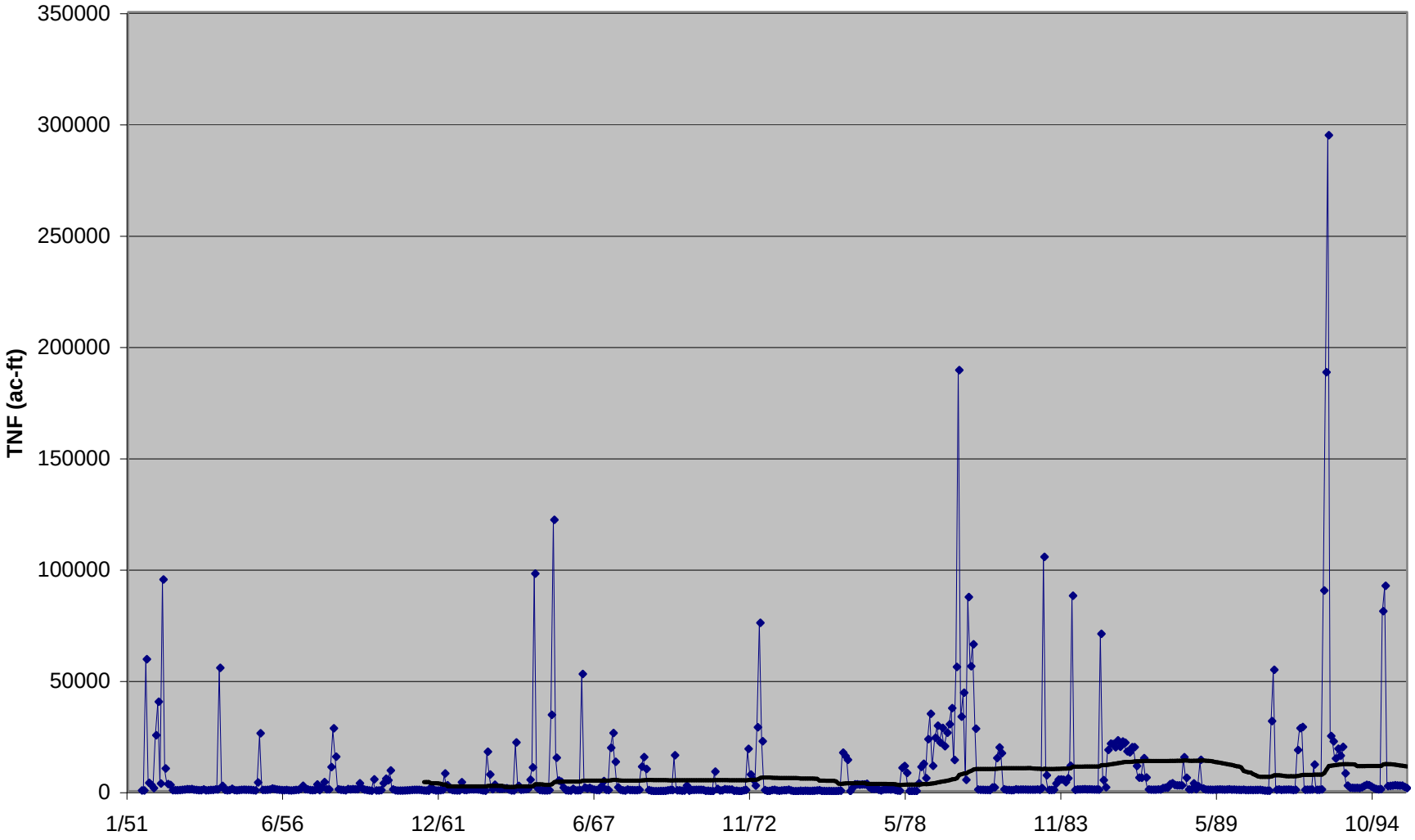
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 25



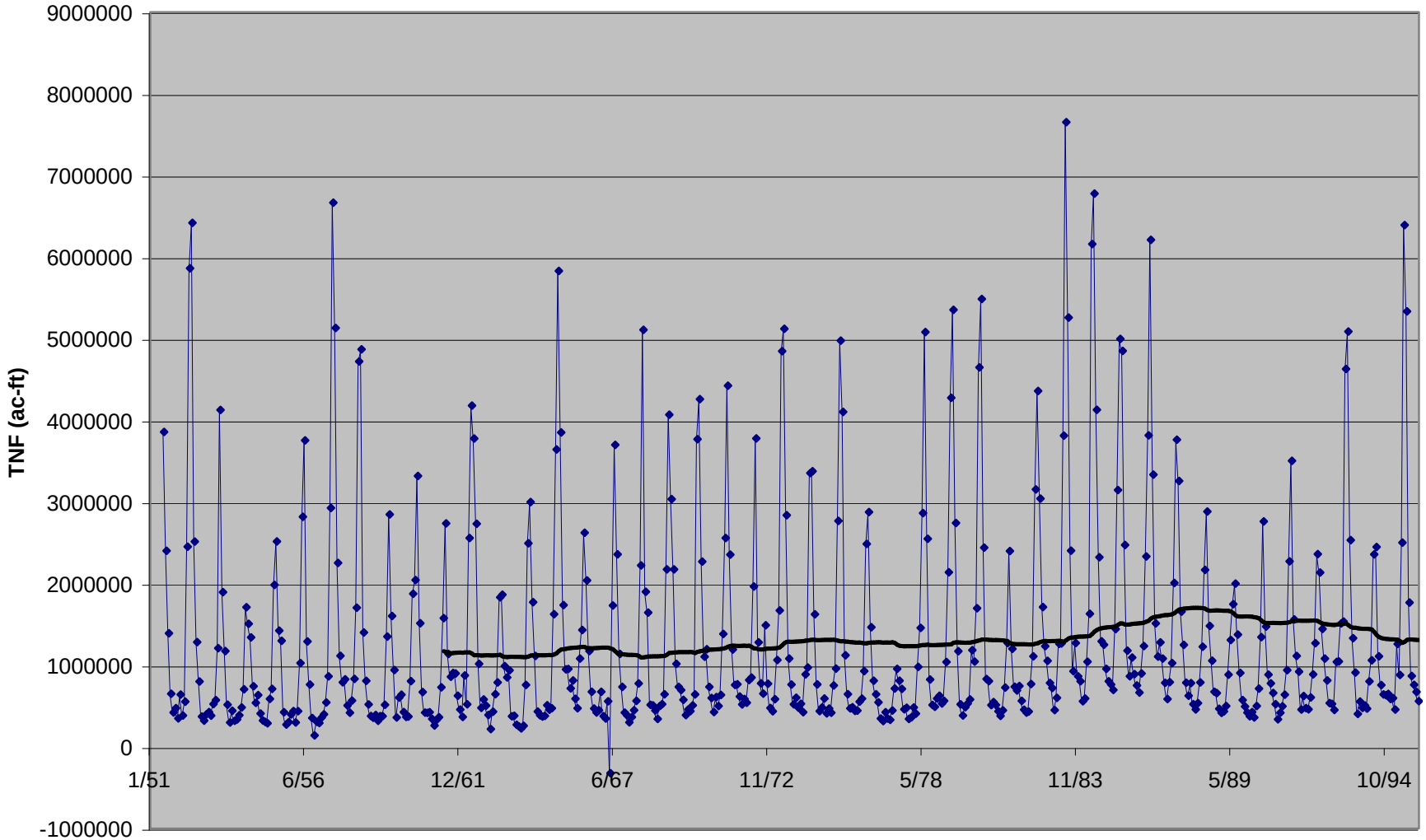
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 26



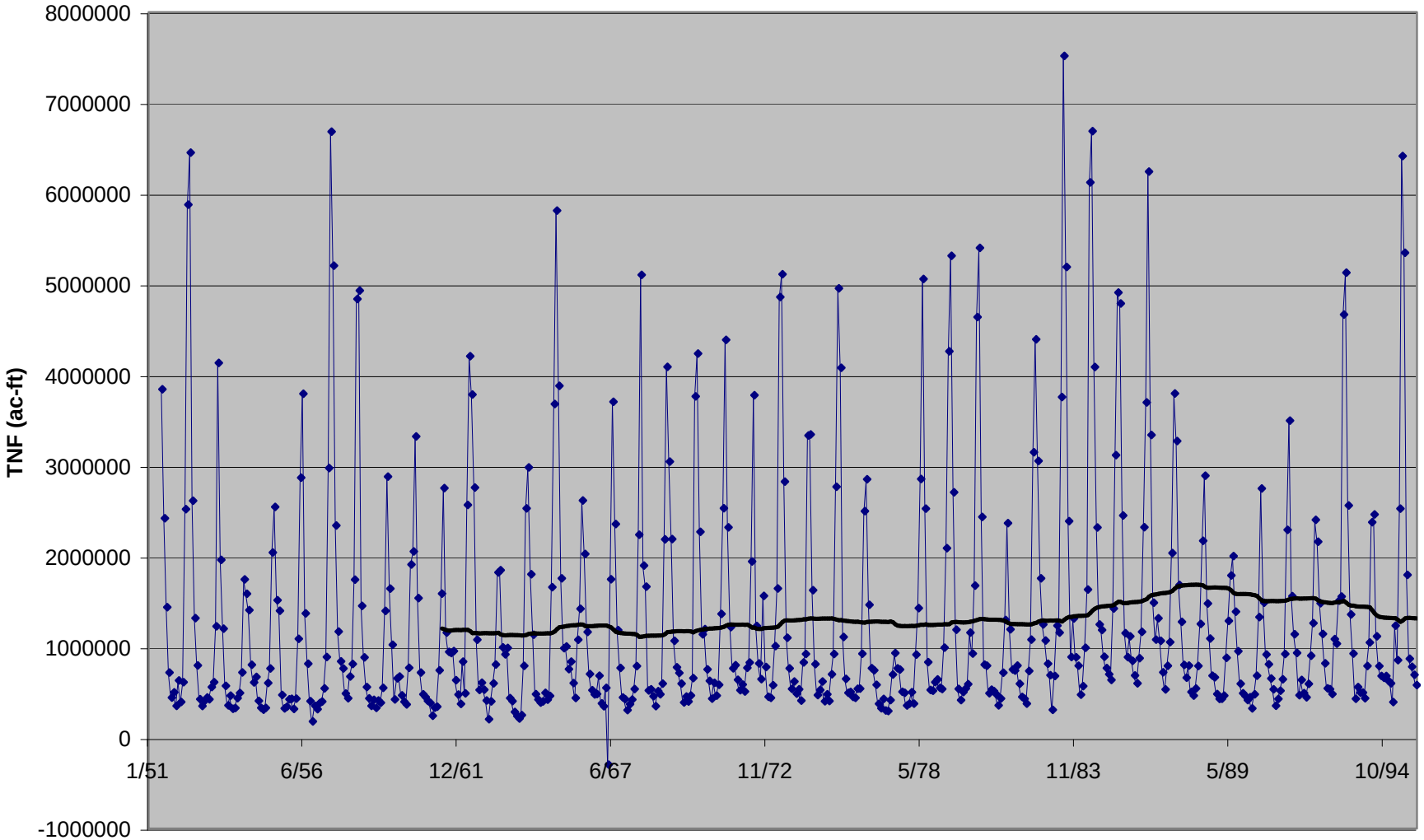
Time Series Data (Total Natural Flow) with 10-yr moving average
Site 27



Time Series Data (Total Natural Flow) with 10-yr moving average
Site 28

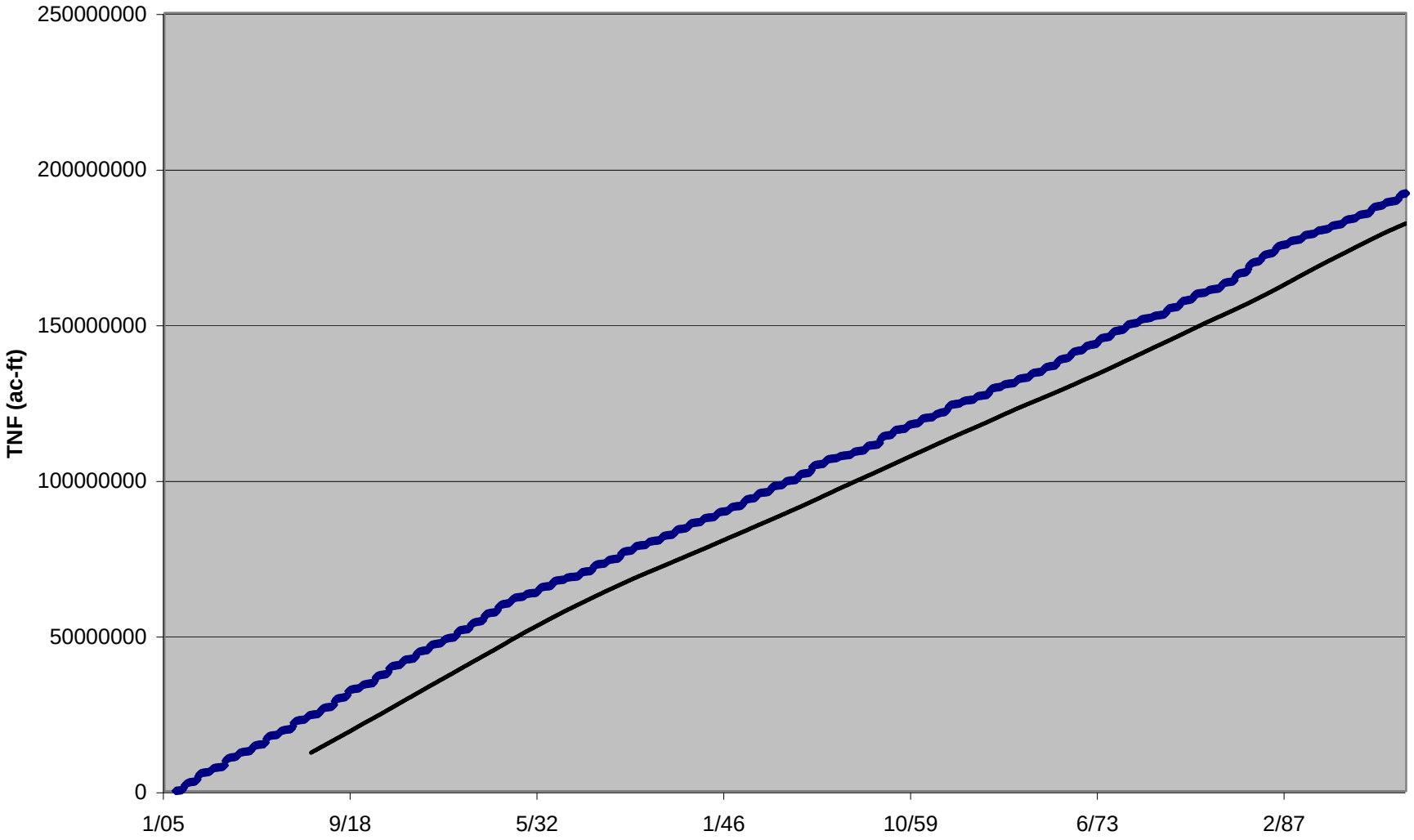


Time Series Data (Total Natural Flow) with 10-yr moving average
Site 29

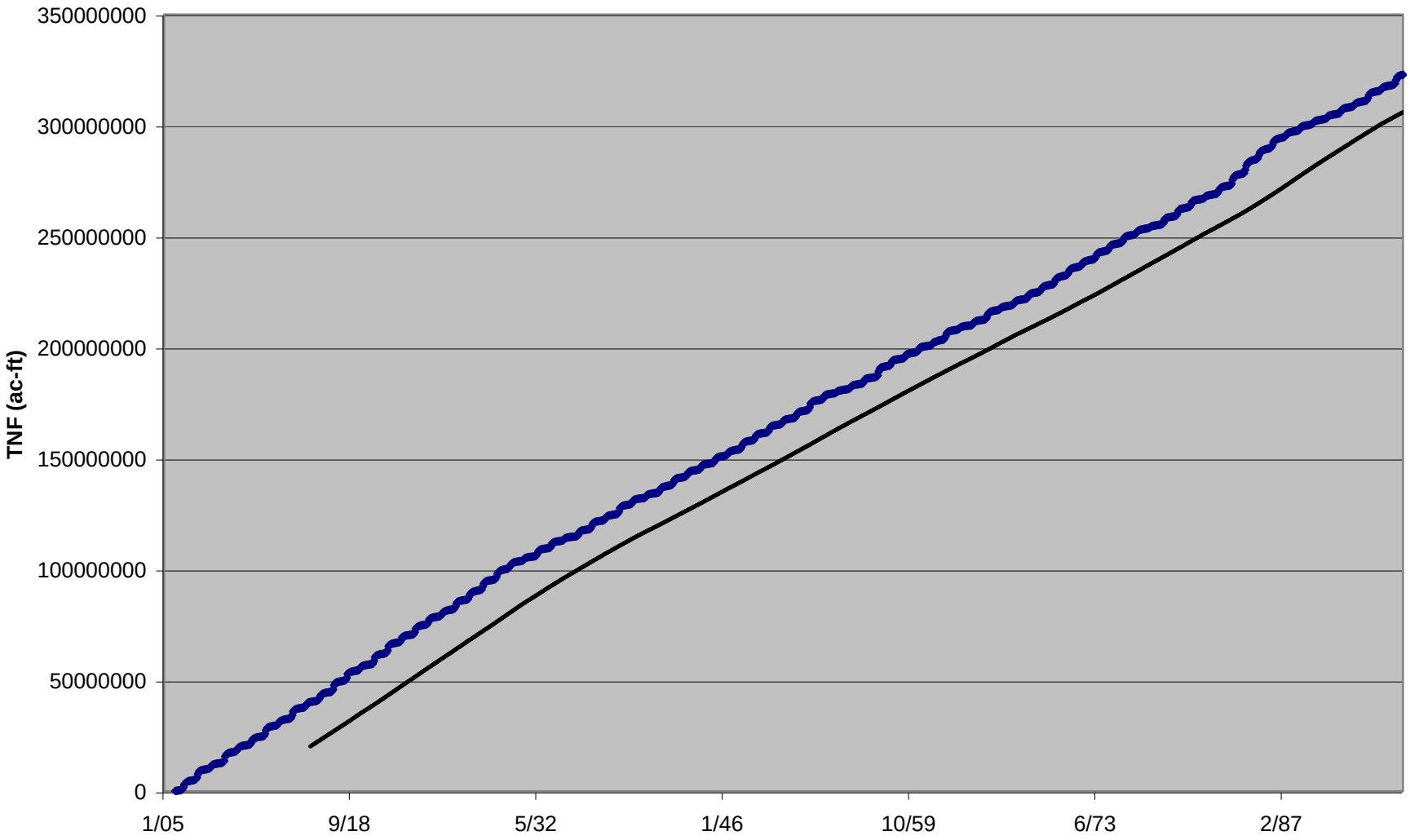


Cumulative Total Natural Flow Plots

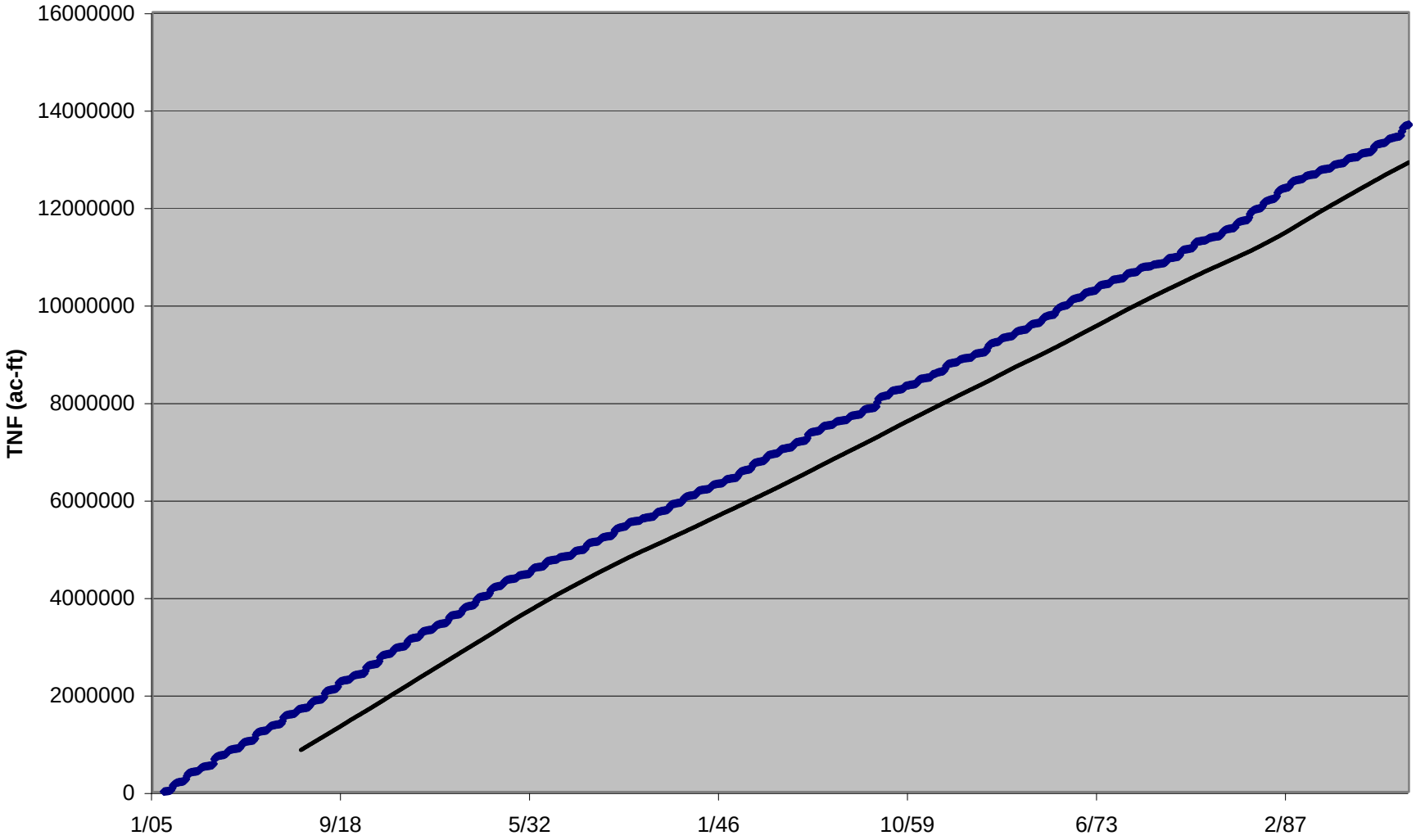
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 1



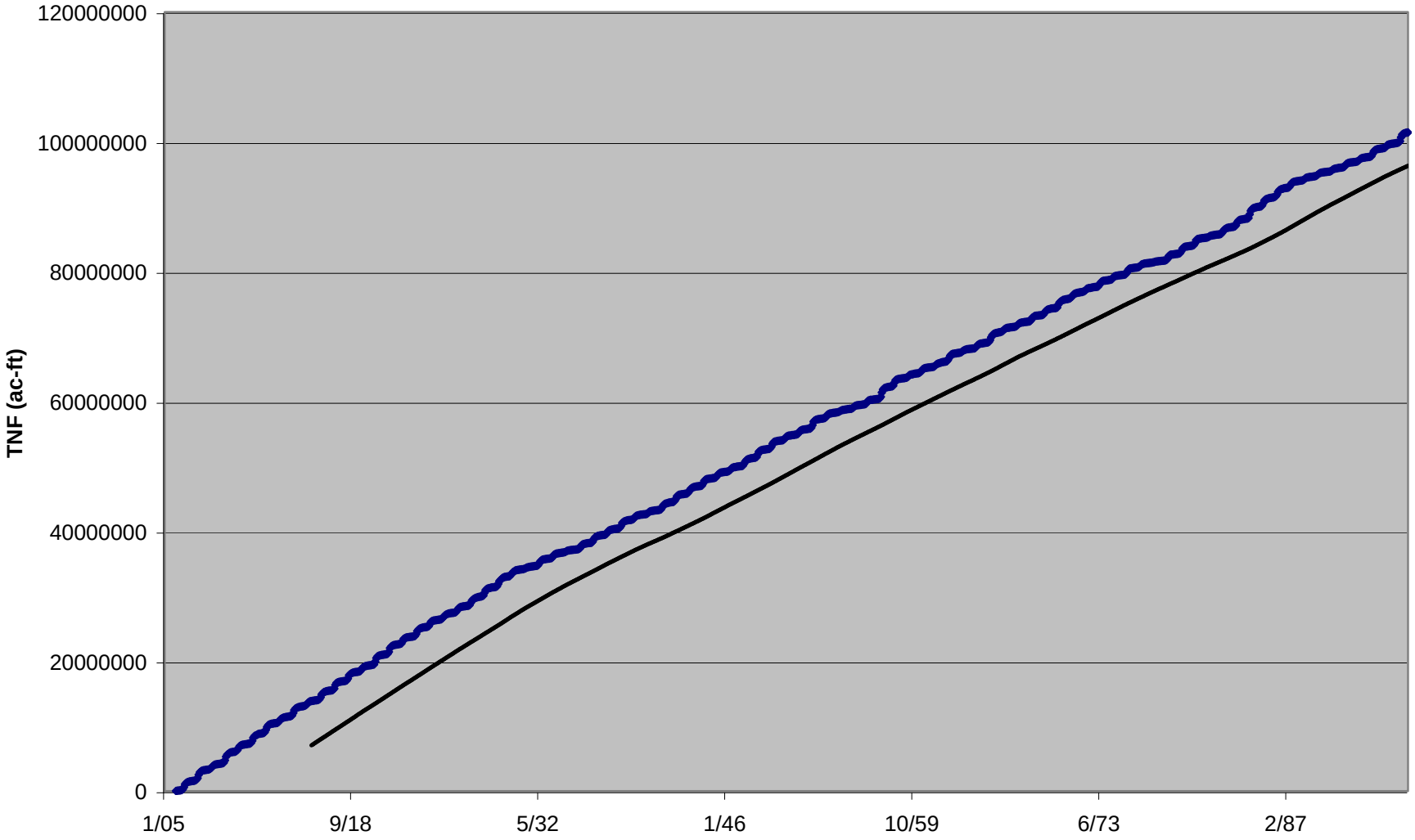
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 2



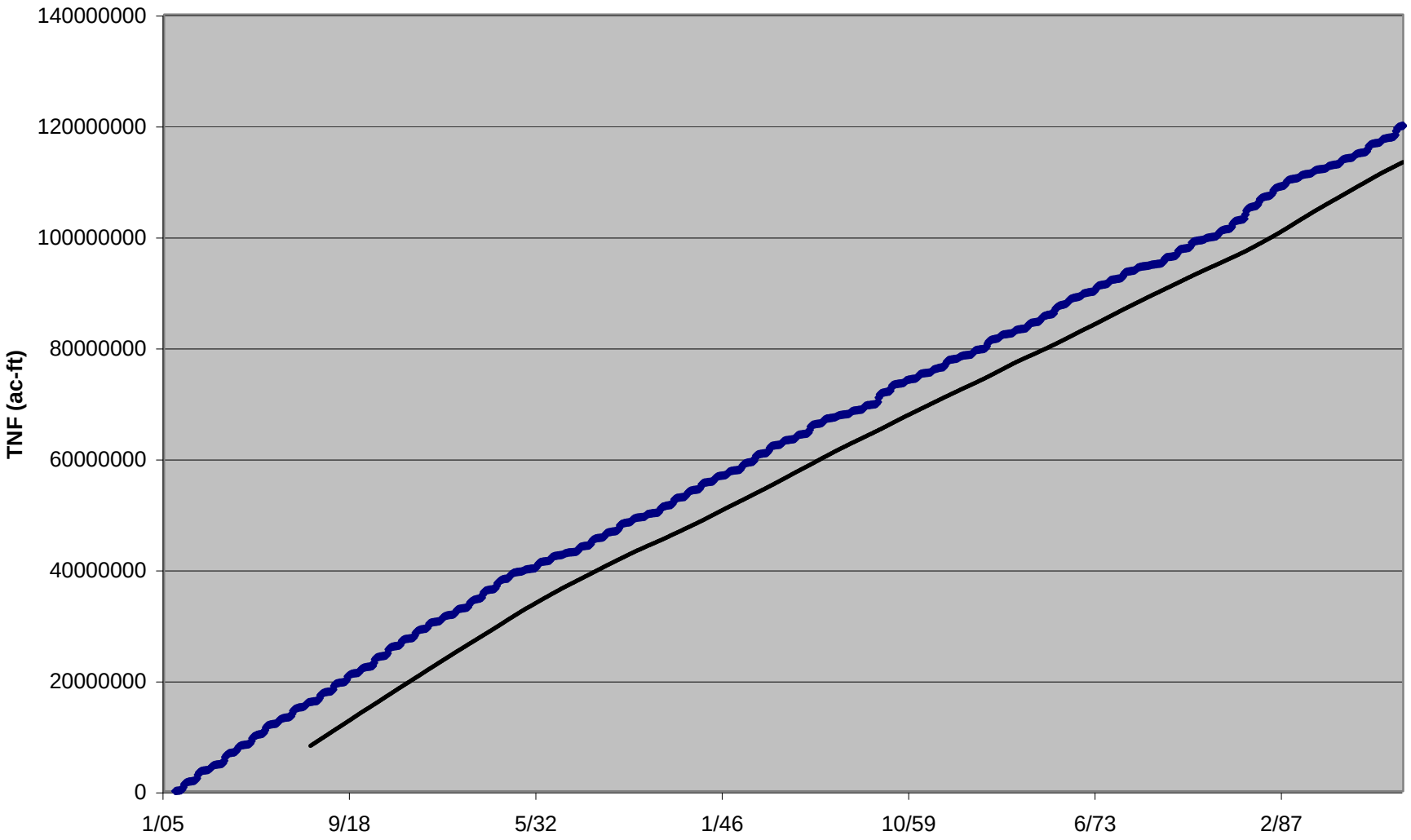
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 3



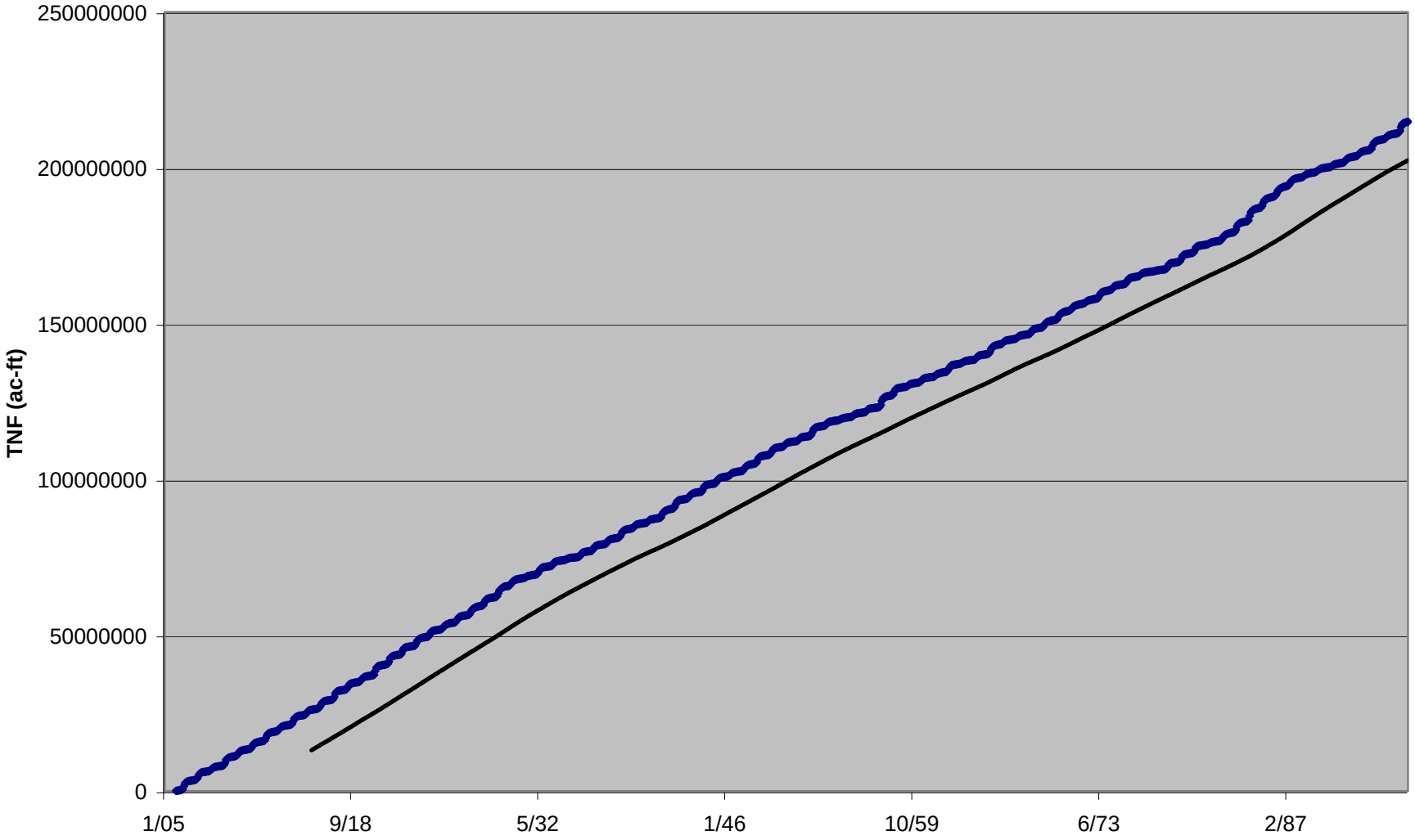
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 4



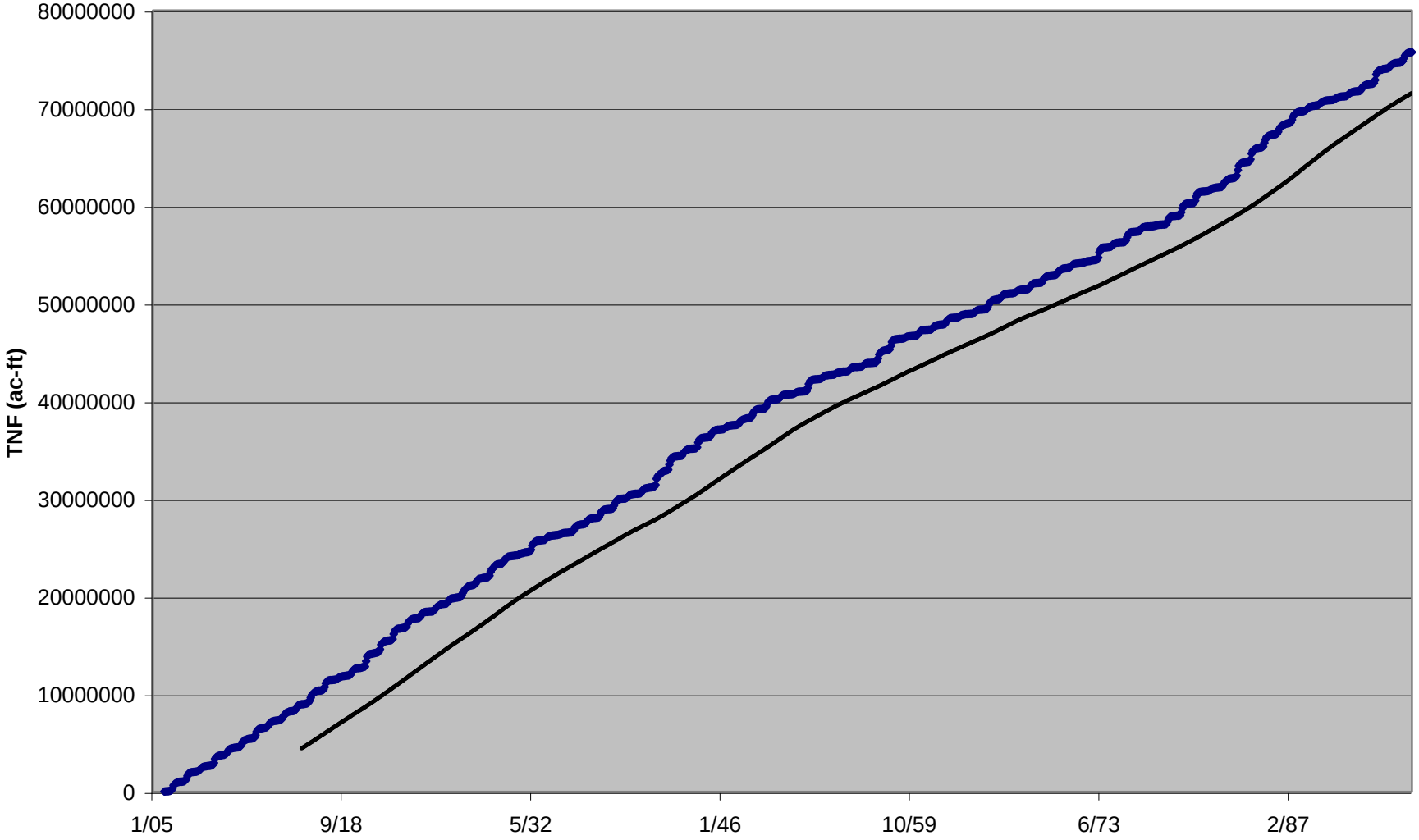
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 5



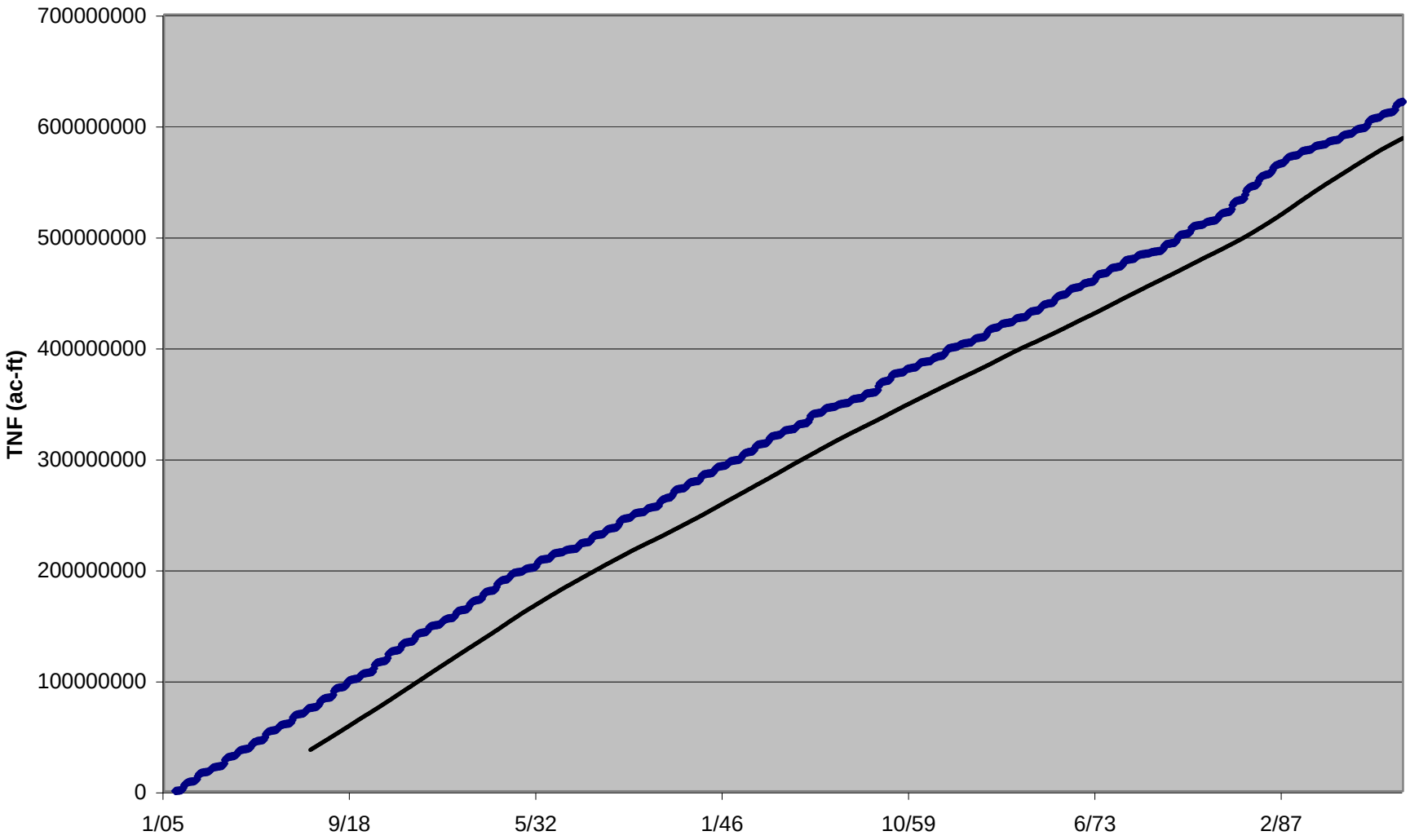
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 6



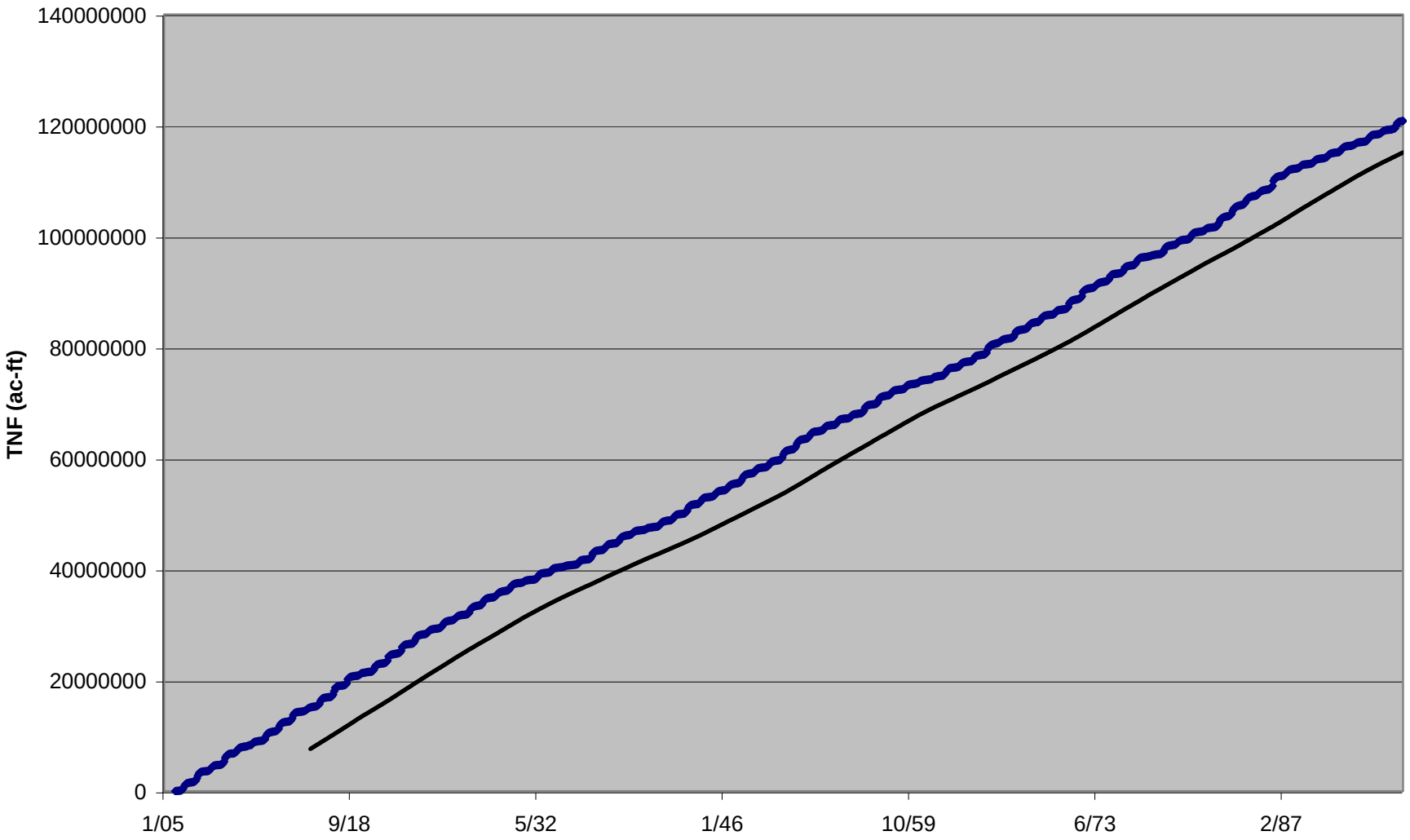
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 7



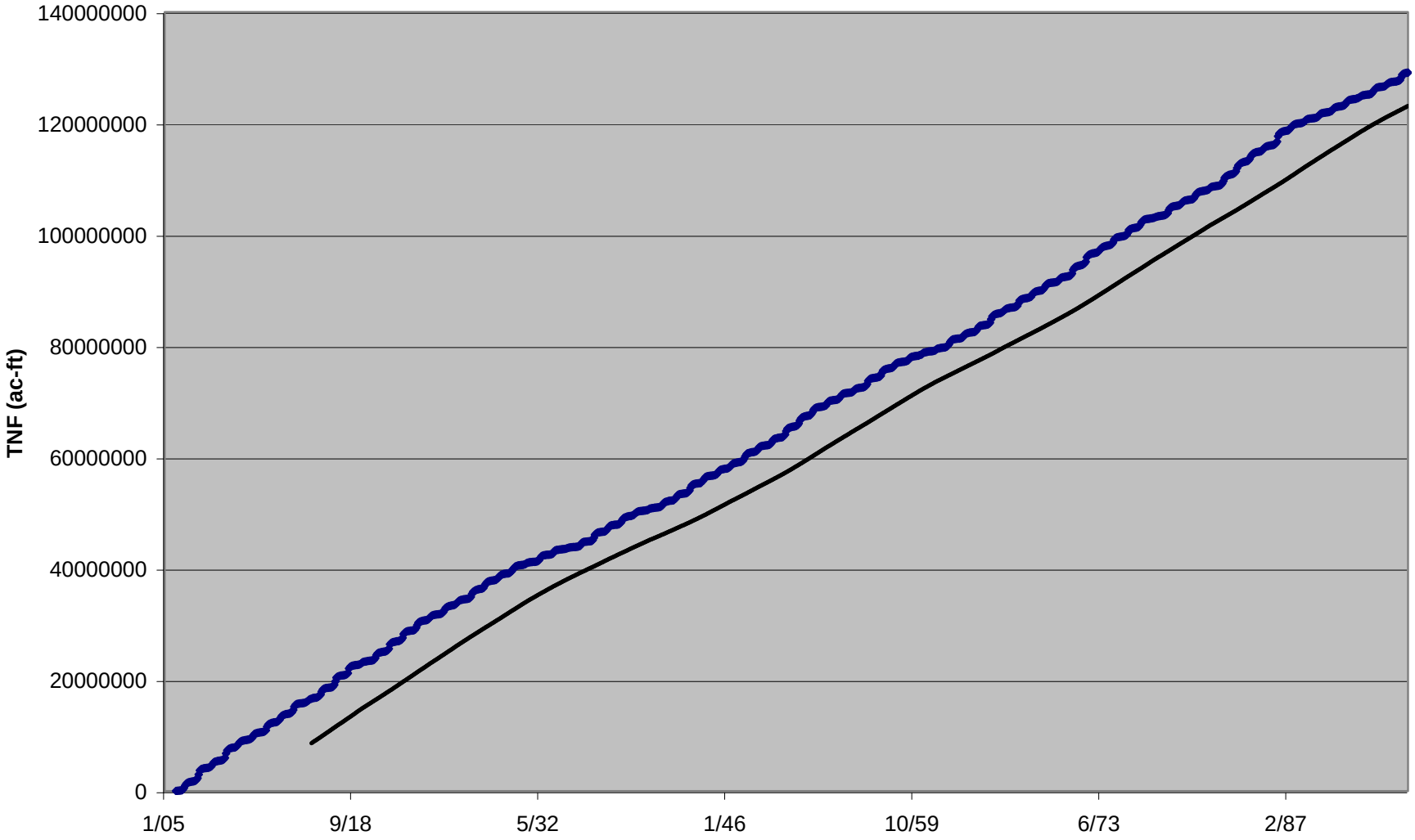
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 8



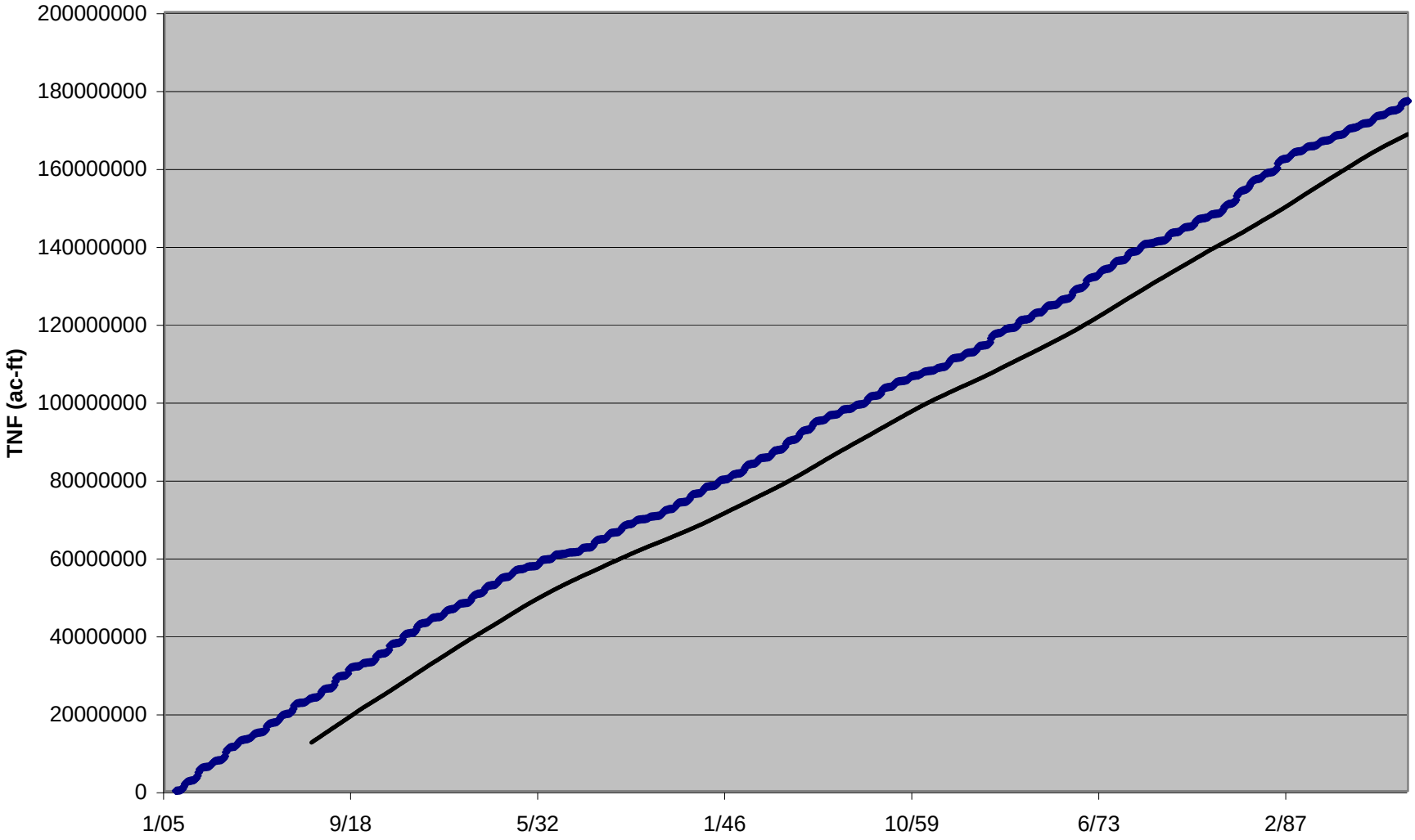
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 9



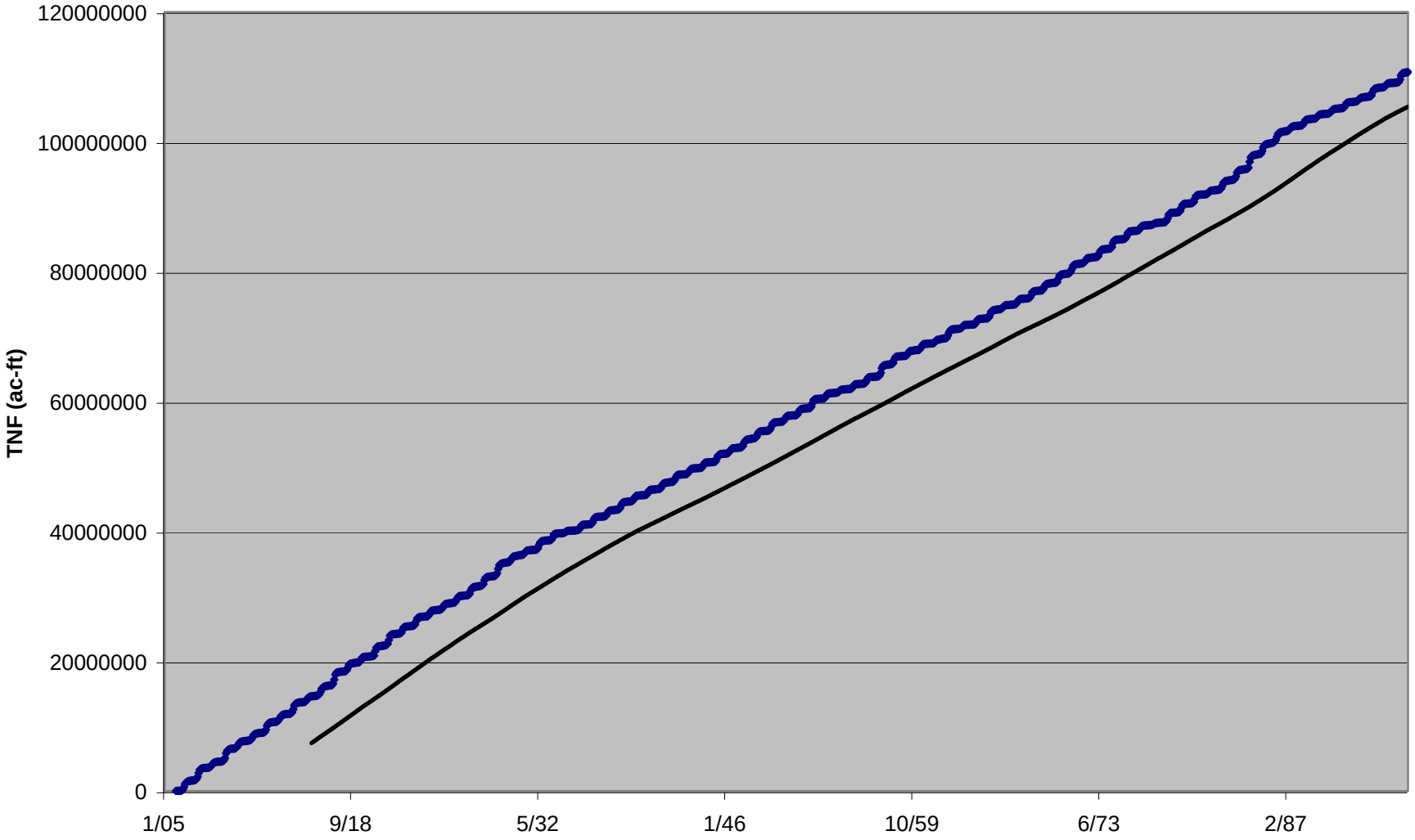
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 10



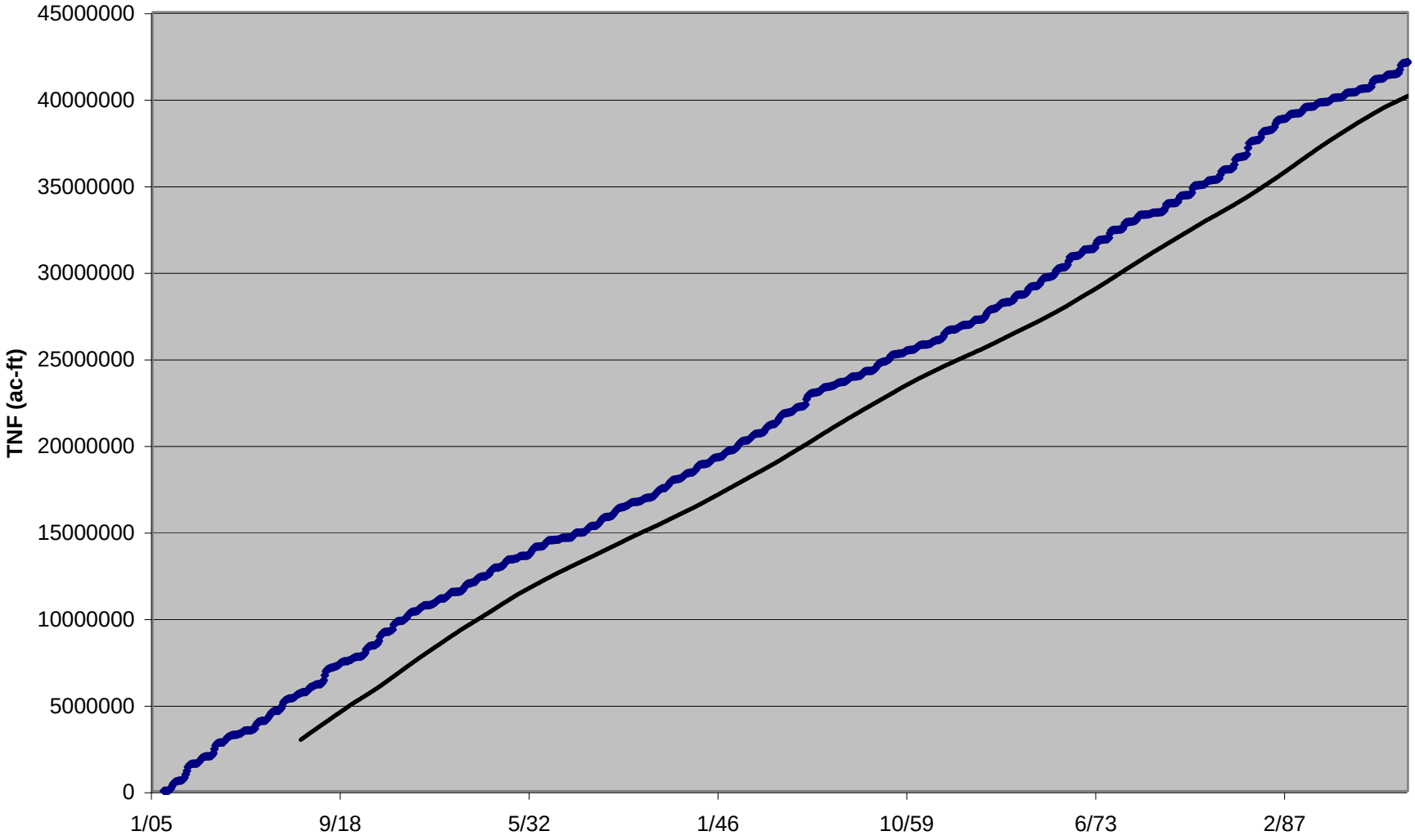
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 11



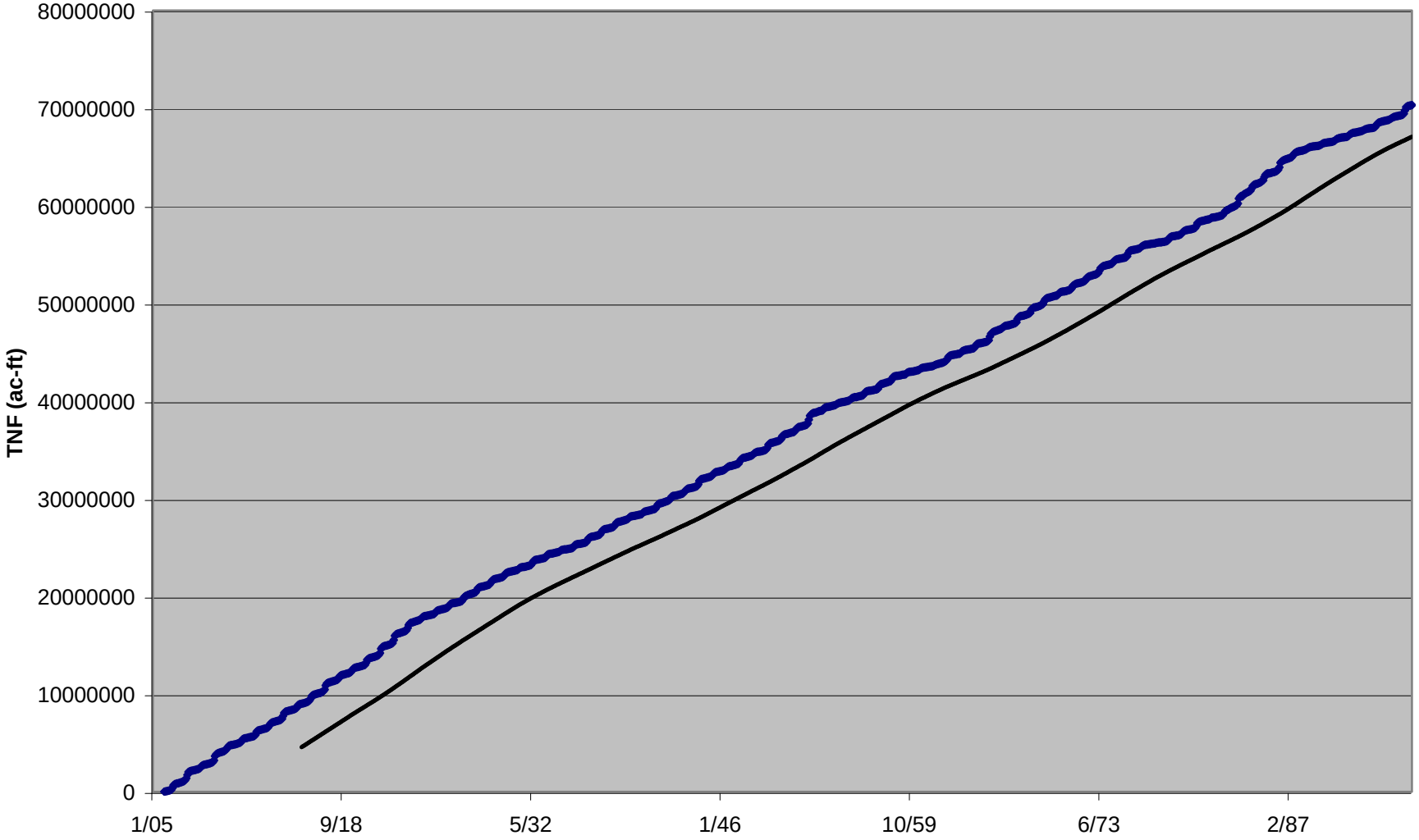
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 12



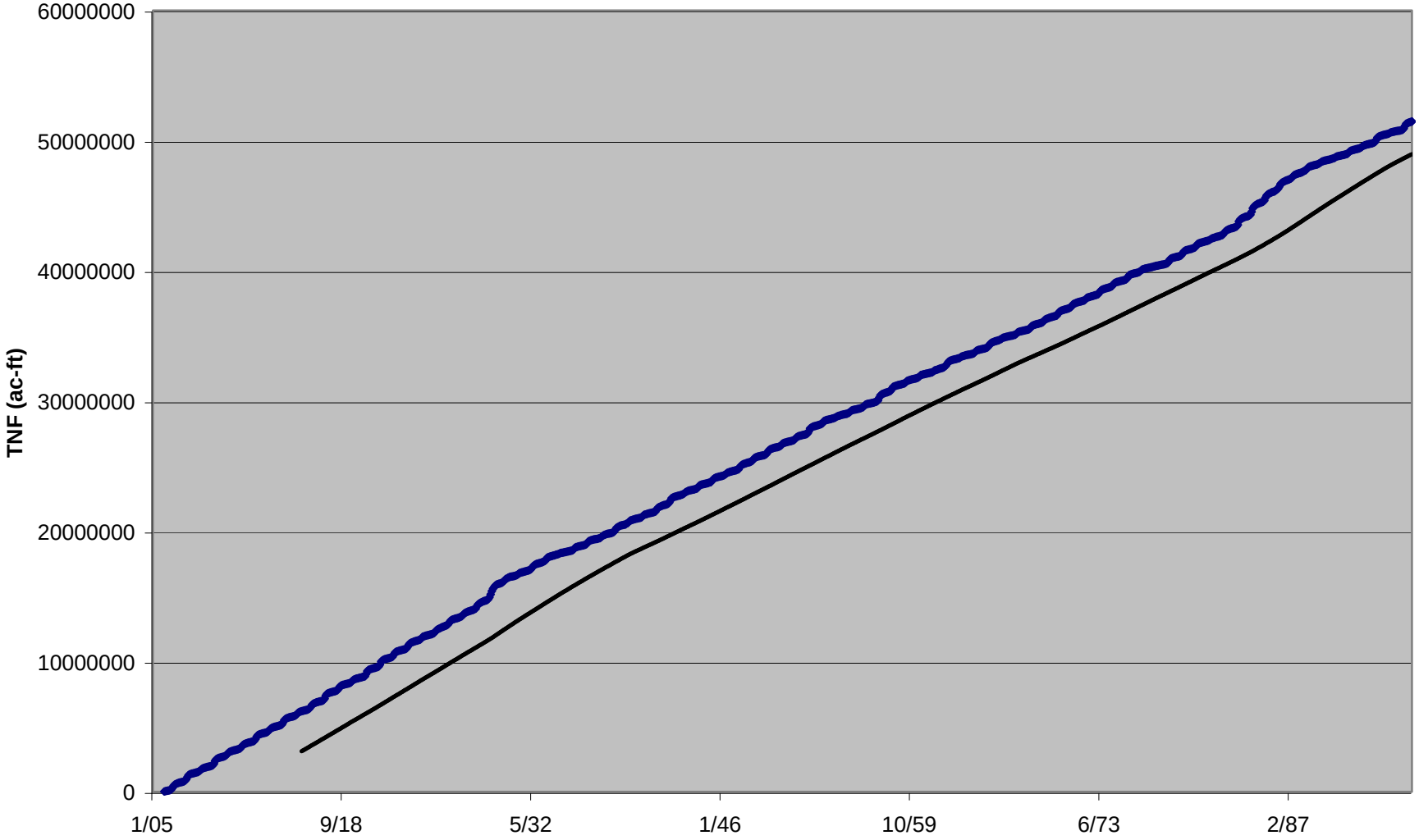
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 13



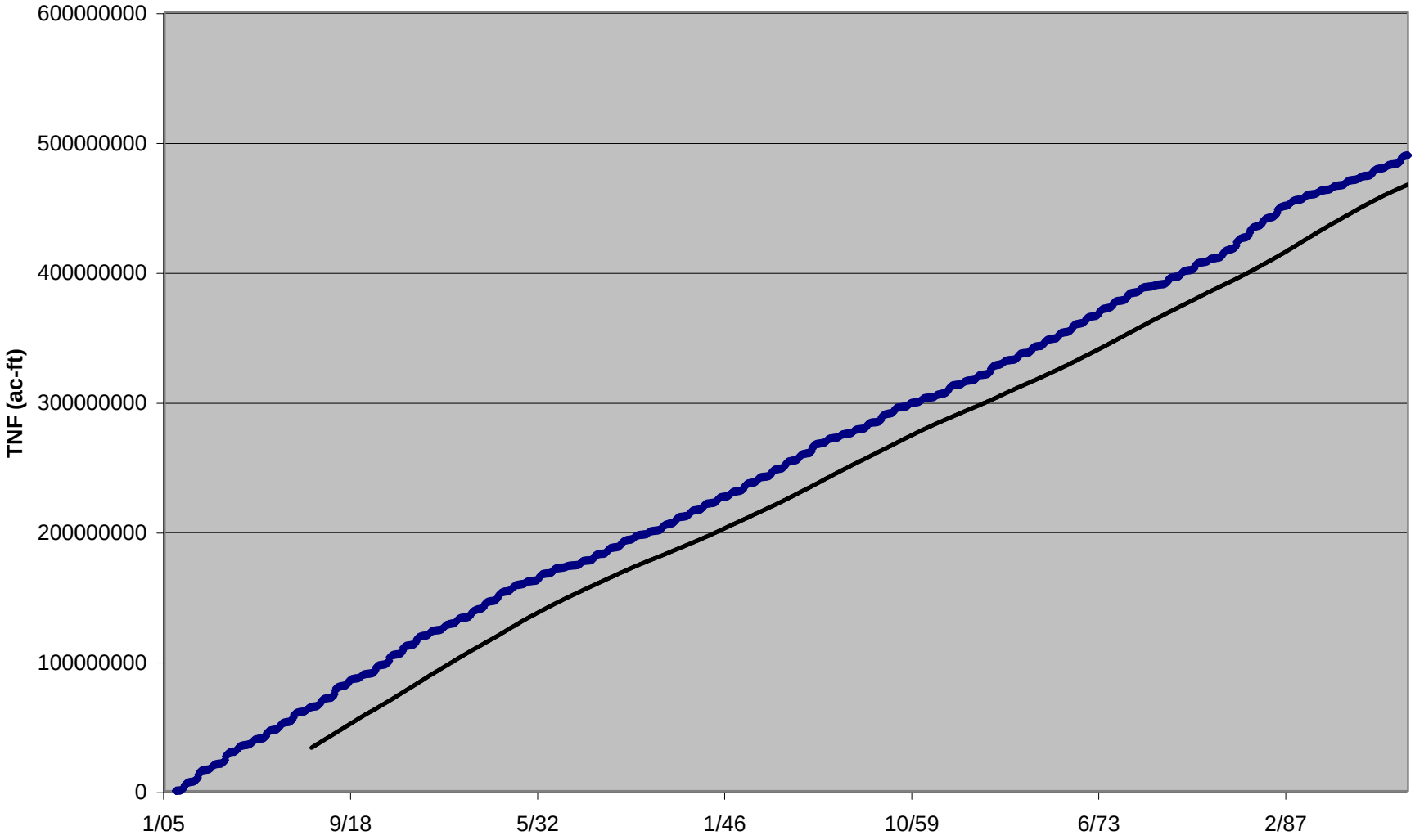
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 14



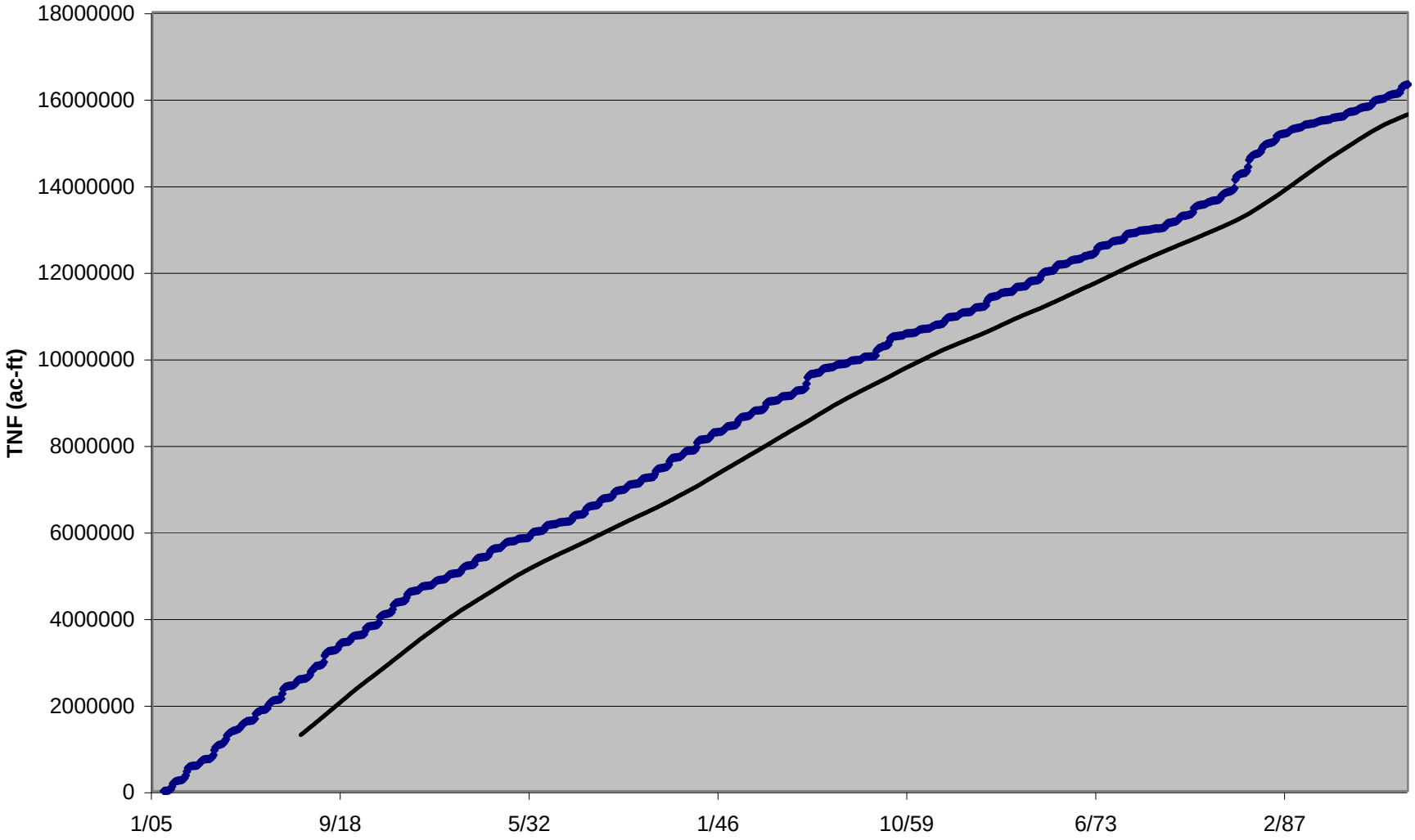
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 15



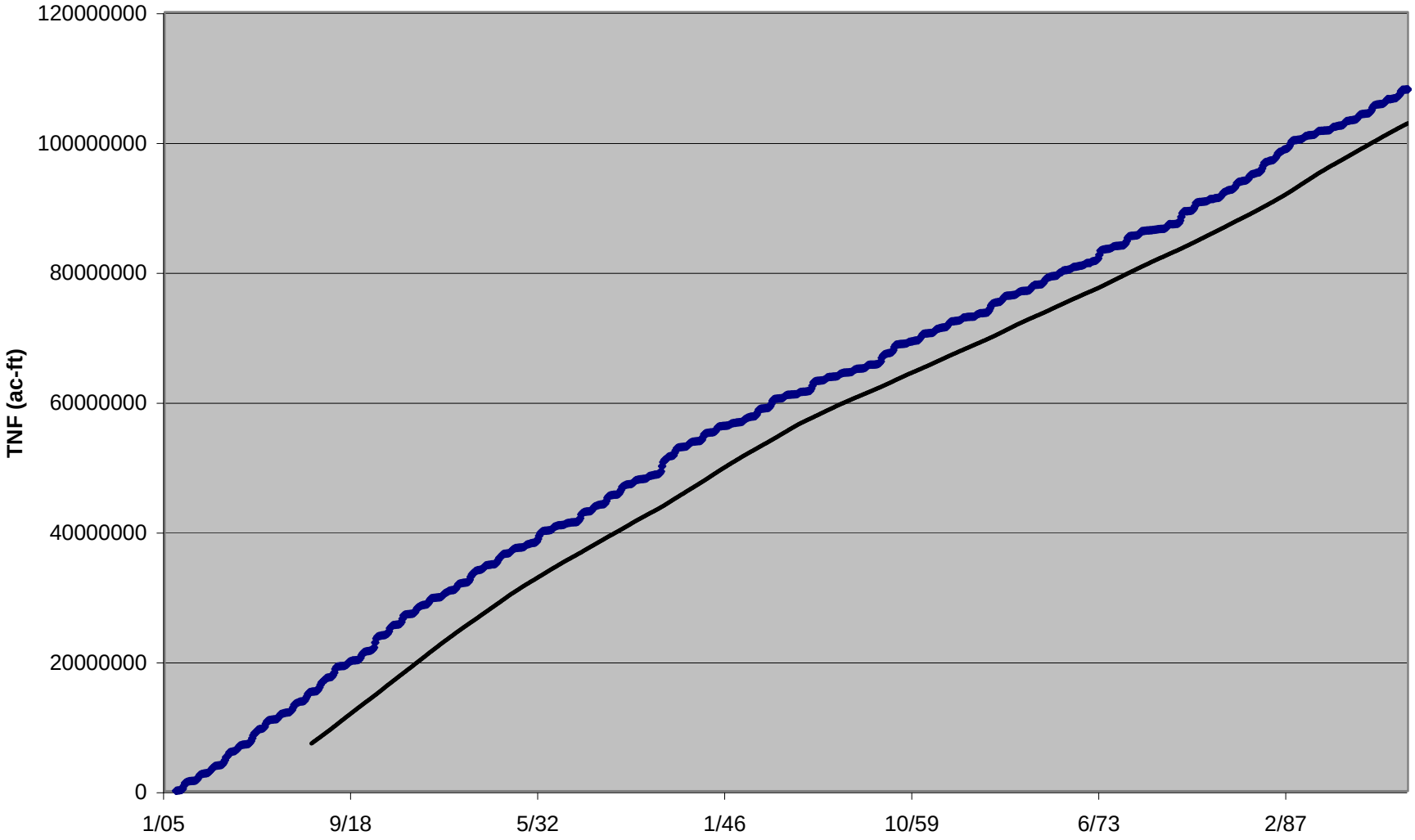
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 16



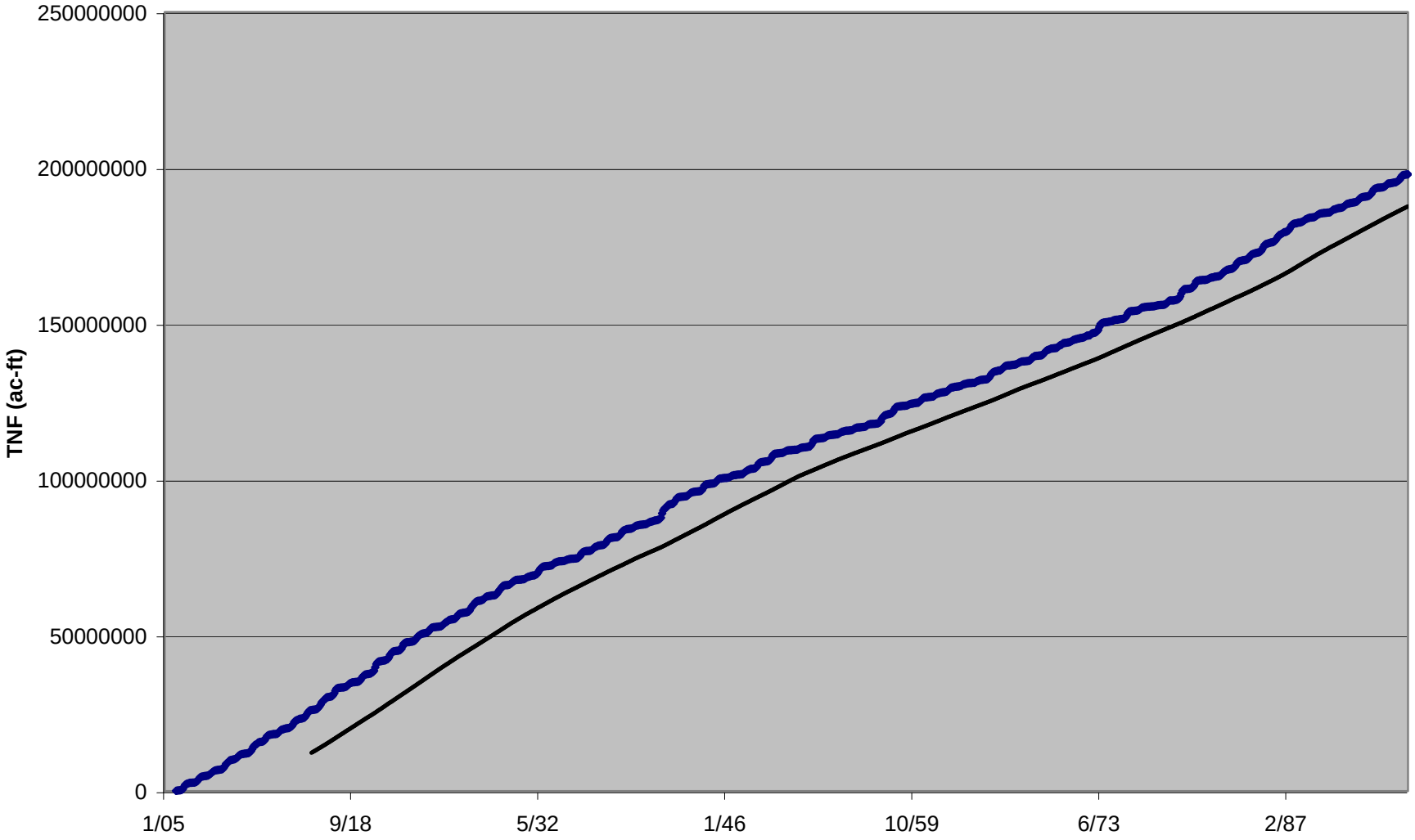
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 17



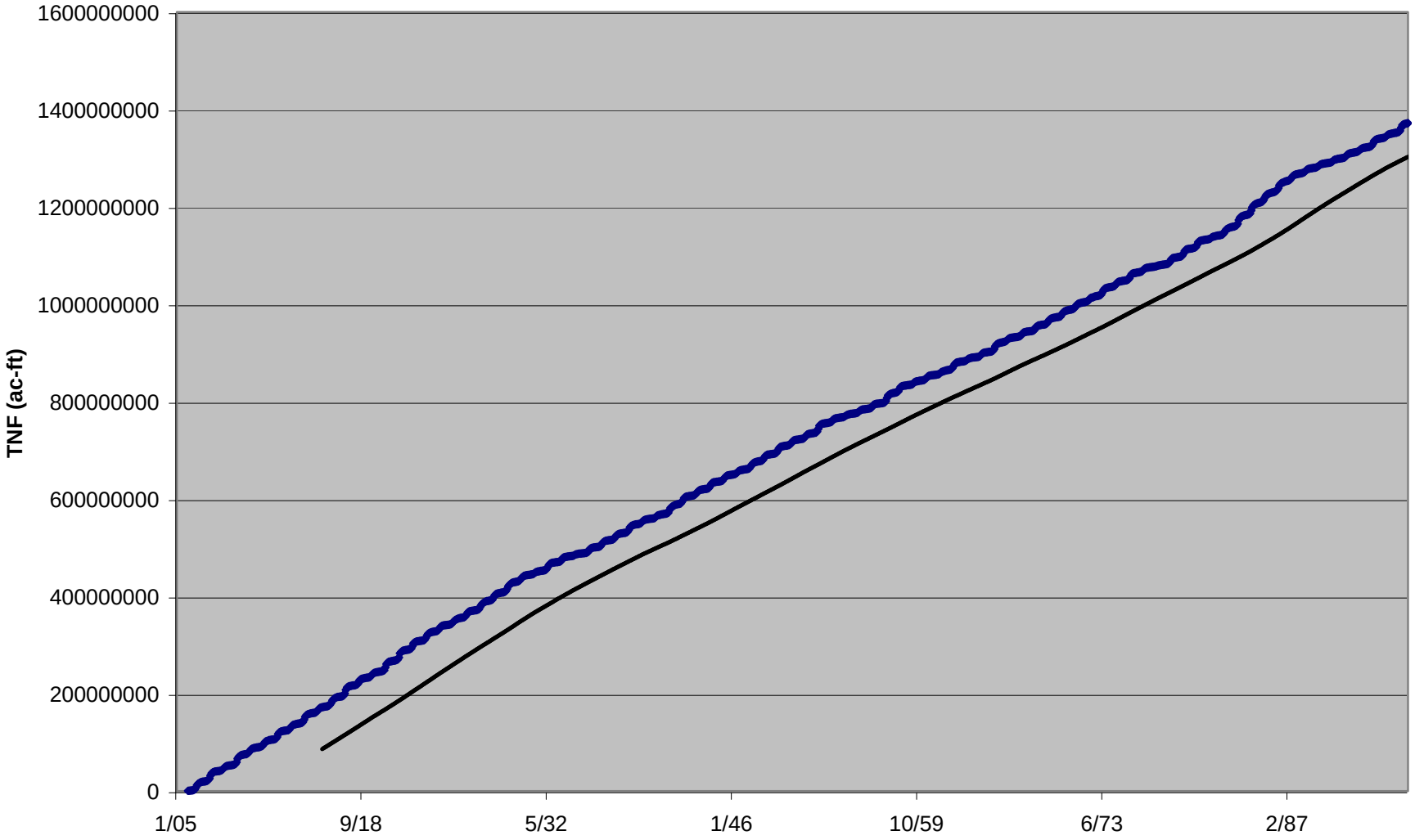
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 18



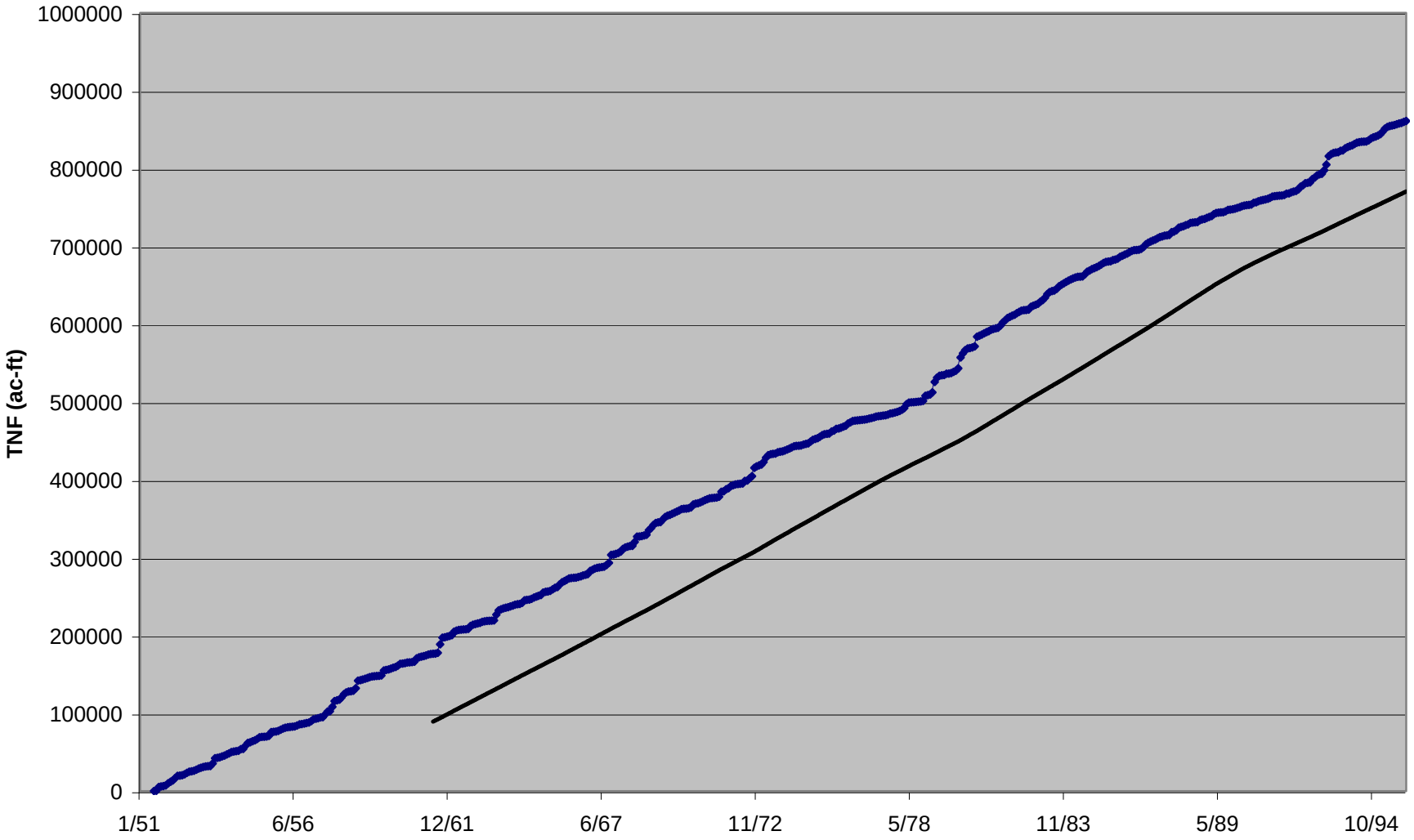
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 19



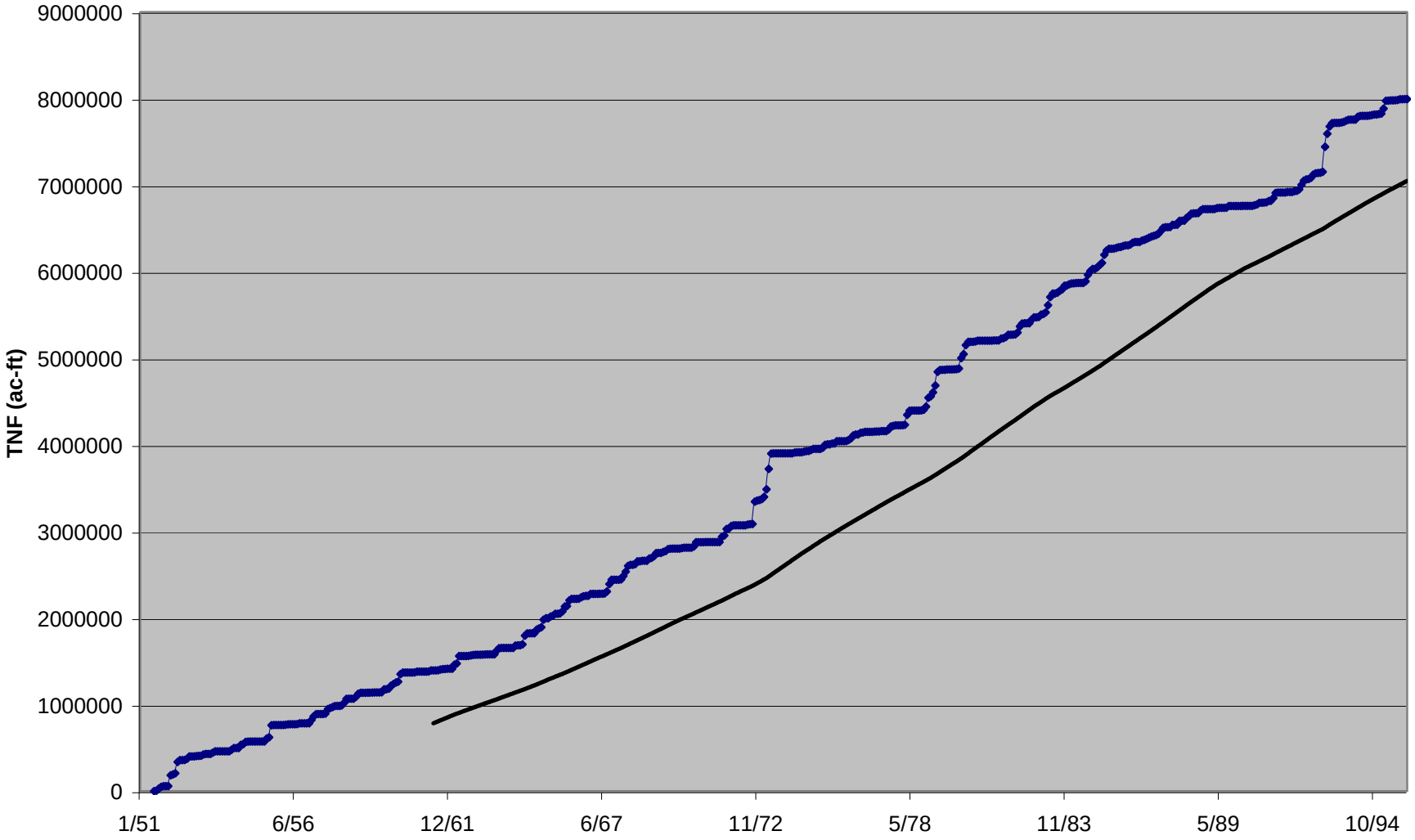
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 20



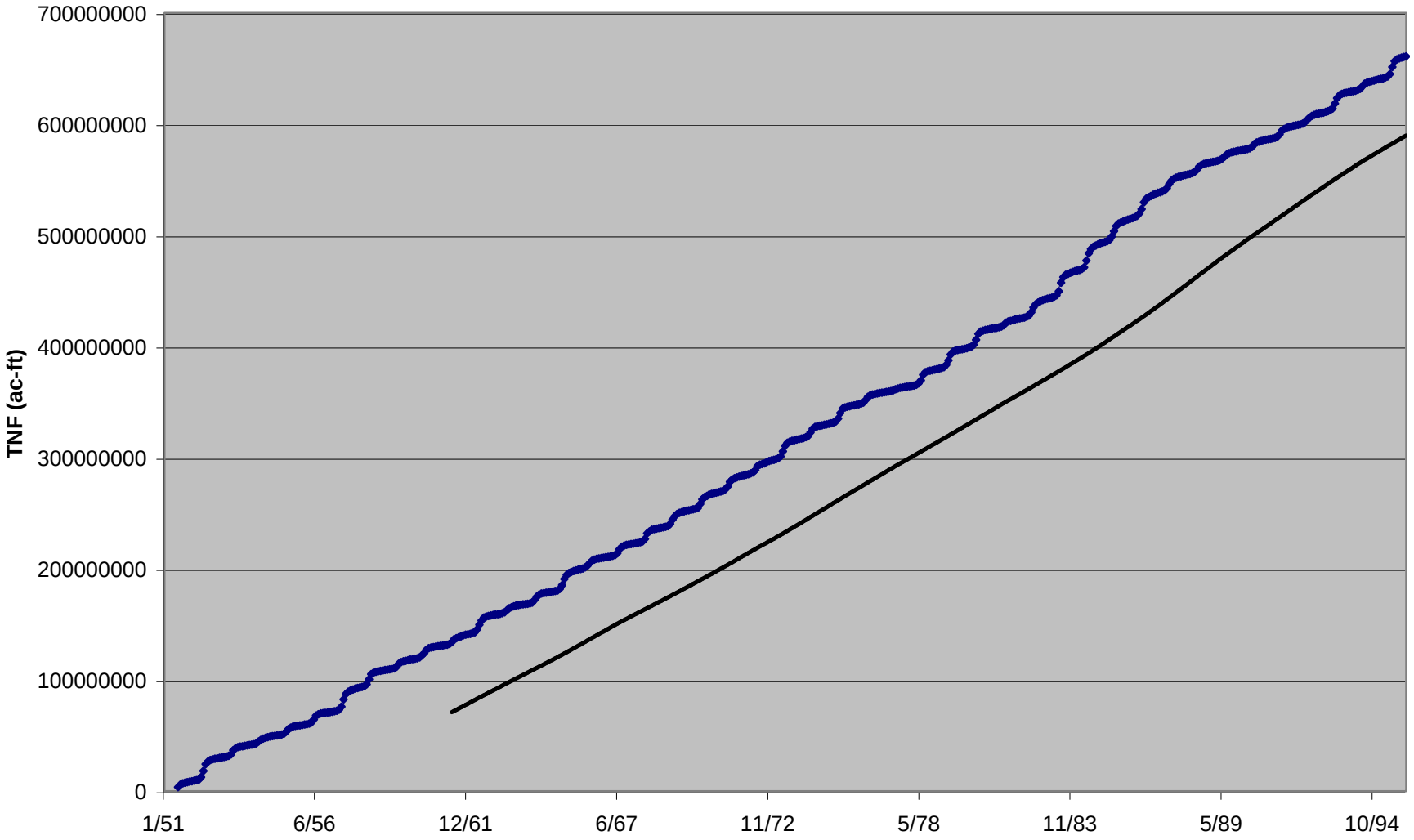
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 21



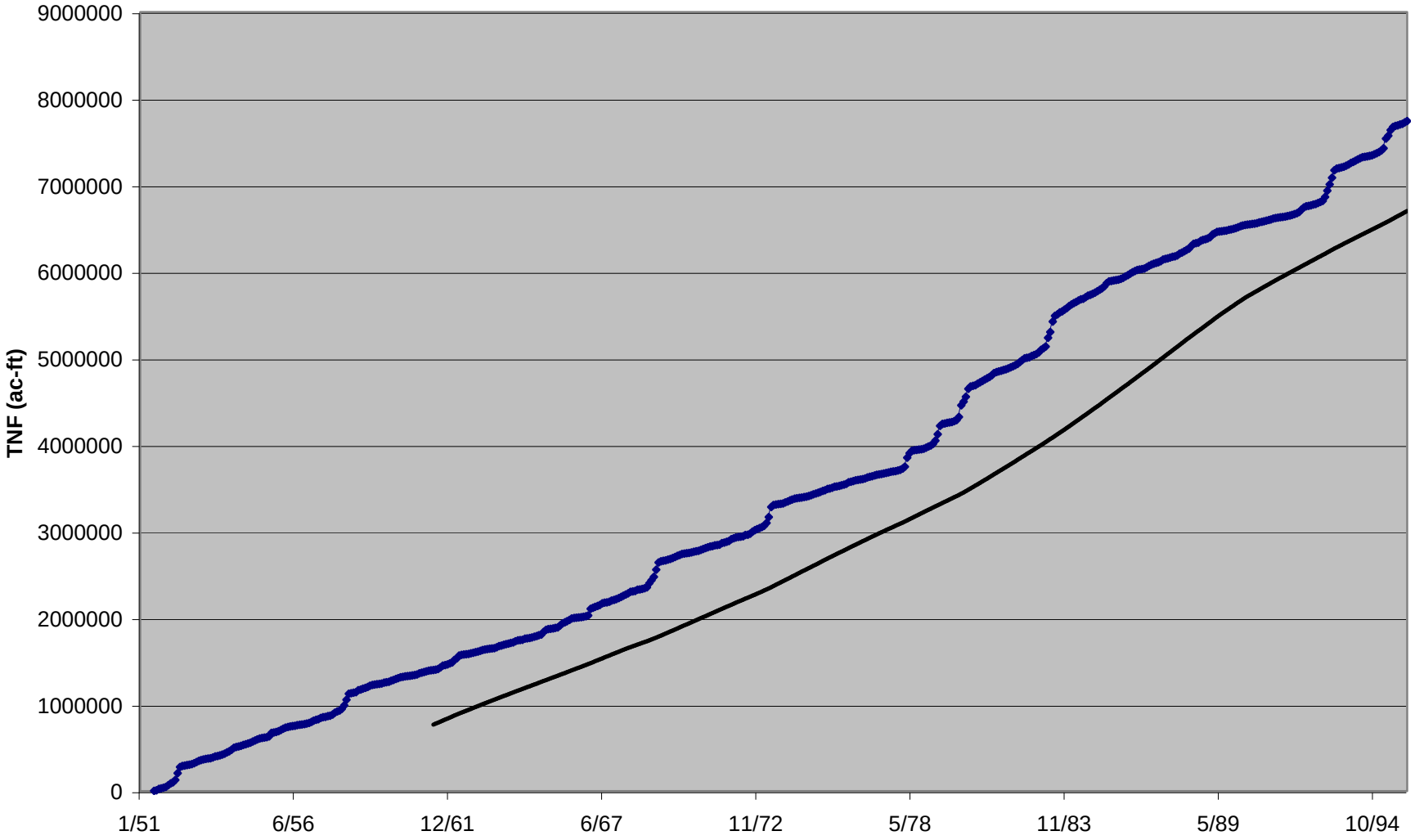
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 22



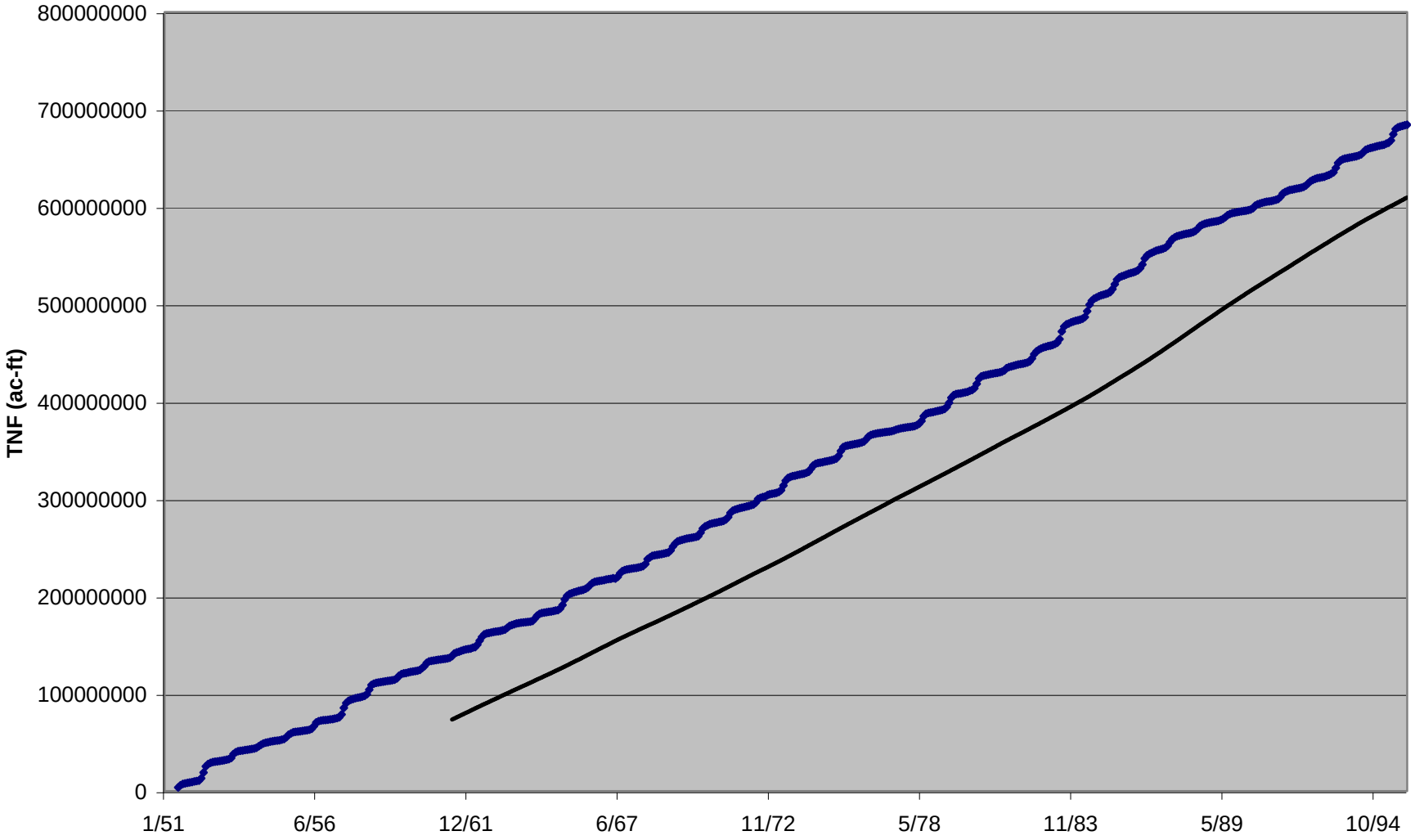
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 23



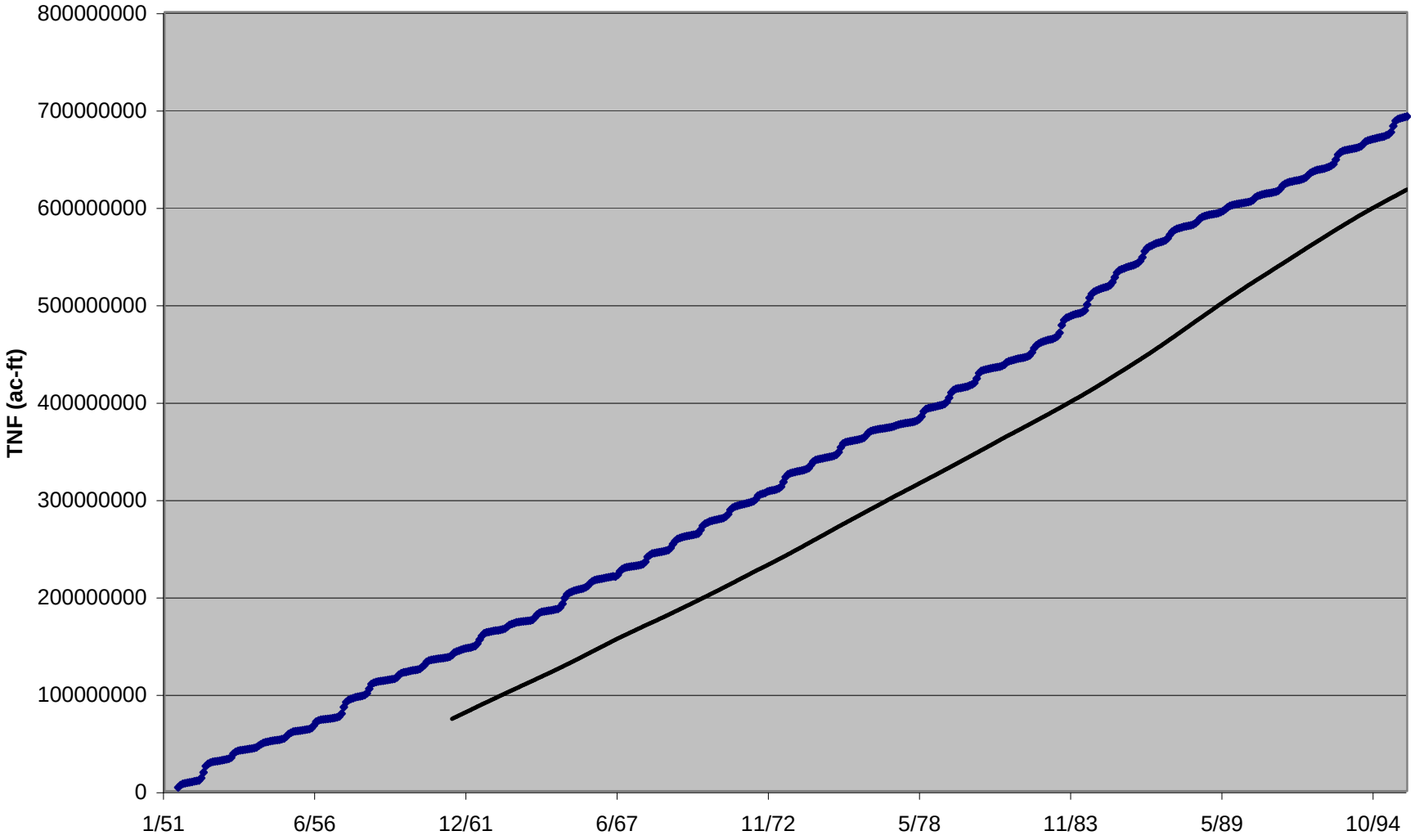
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 24



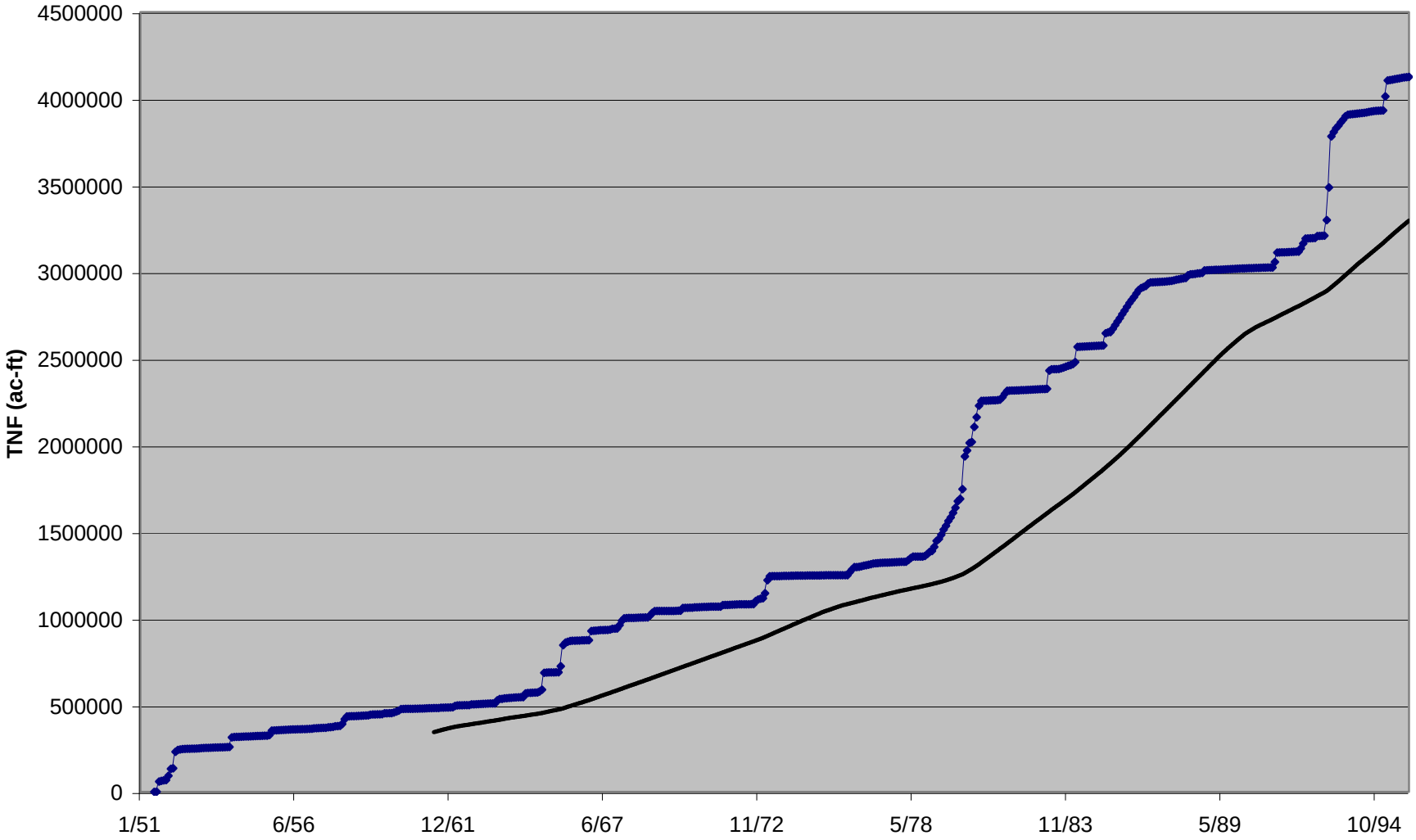
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 25



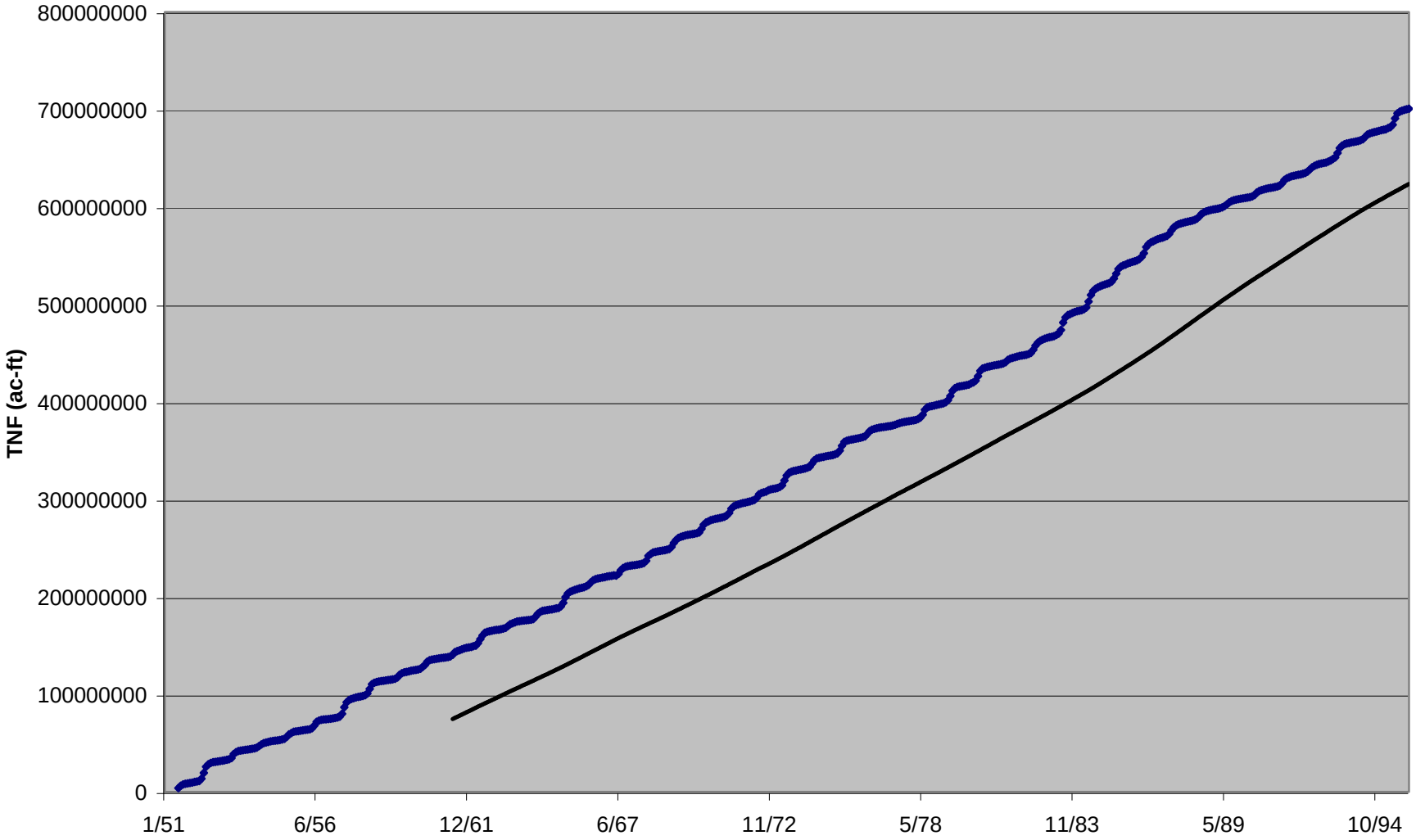
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 26



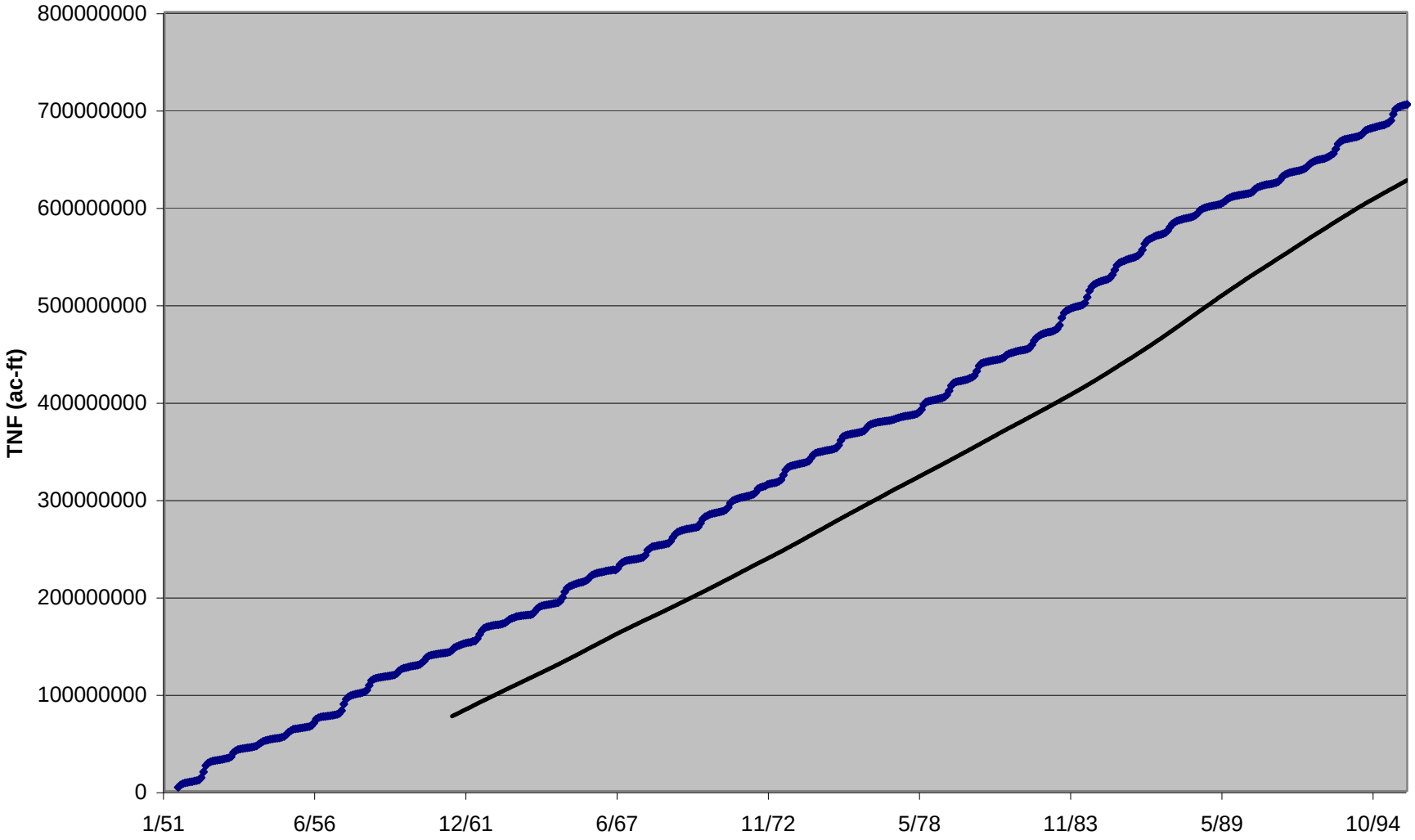
Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 27



Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 28

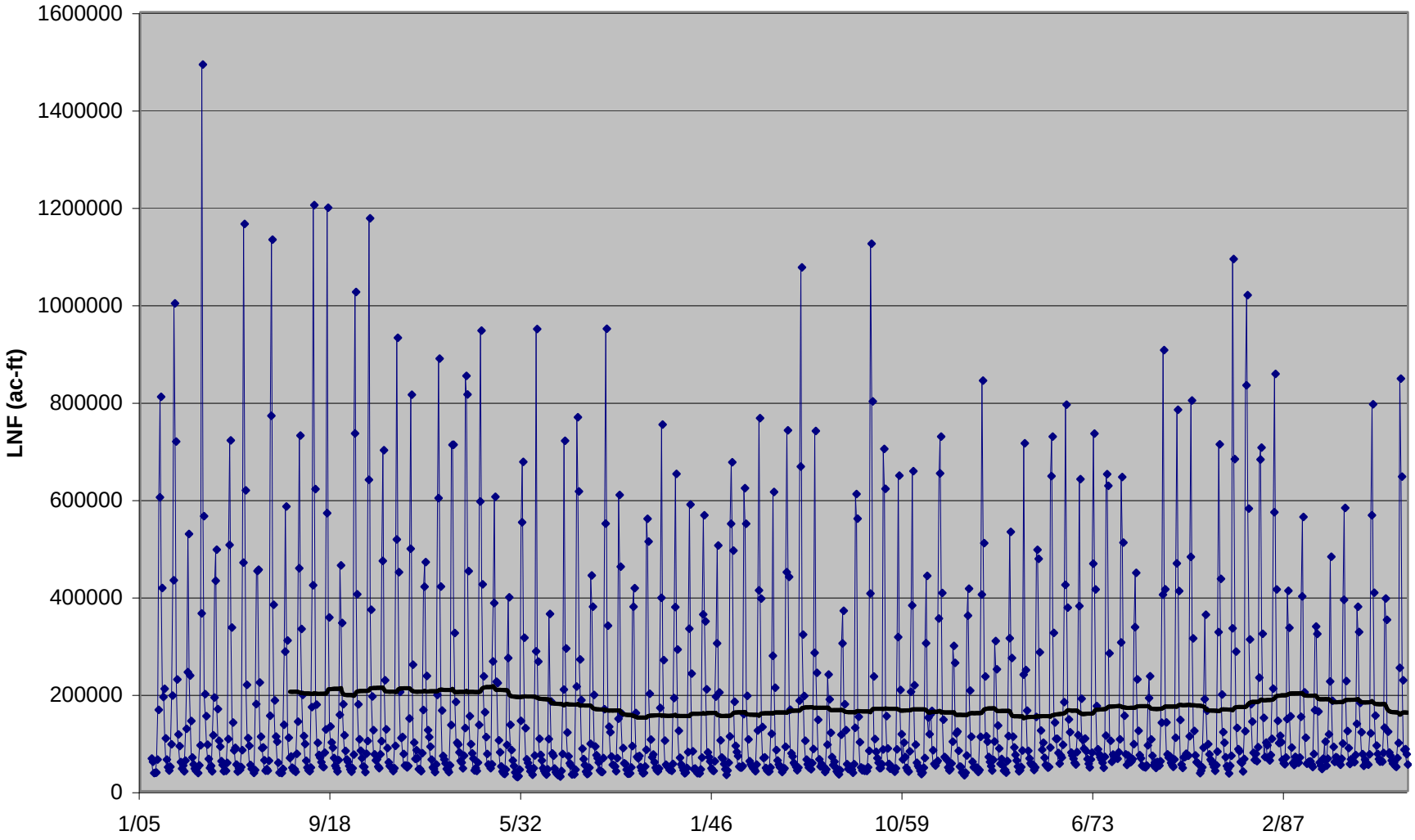


Simple Mass Curve (Total Natural Flow with 10-yr moving average)
Site 29

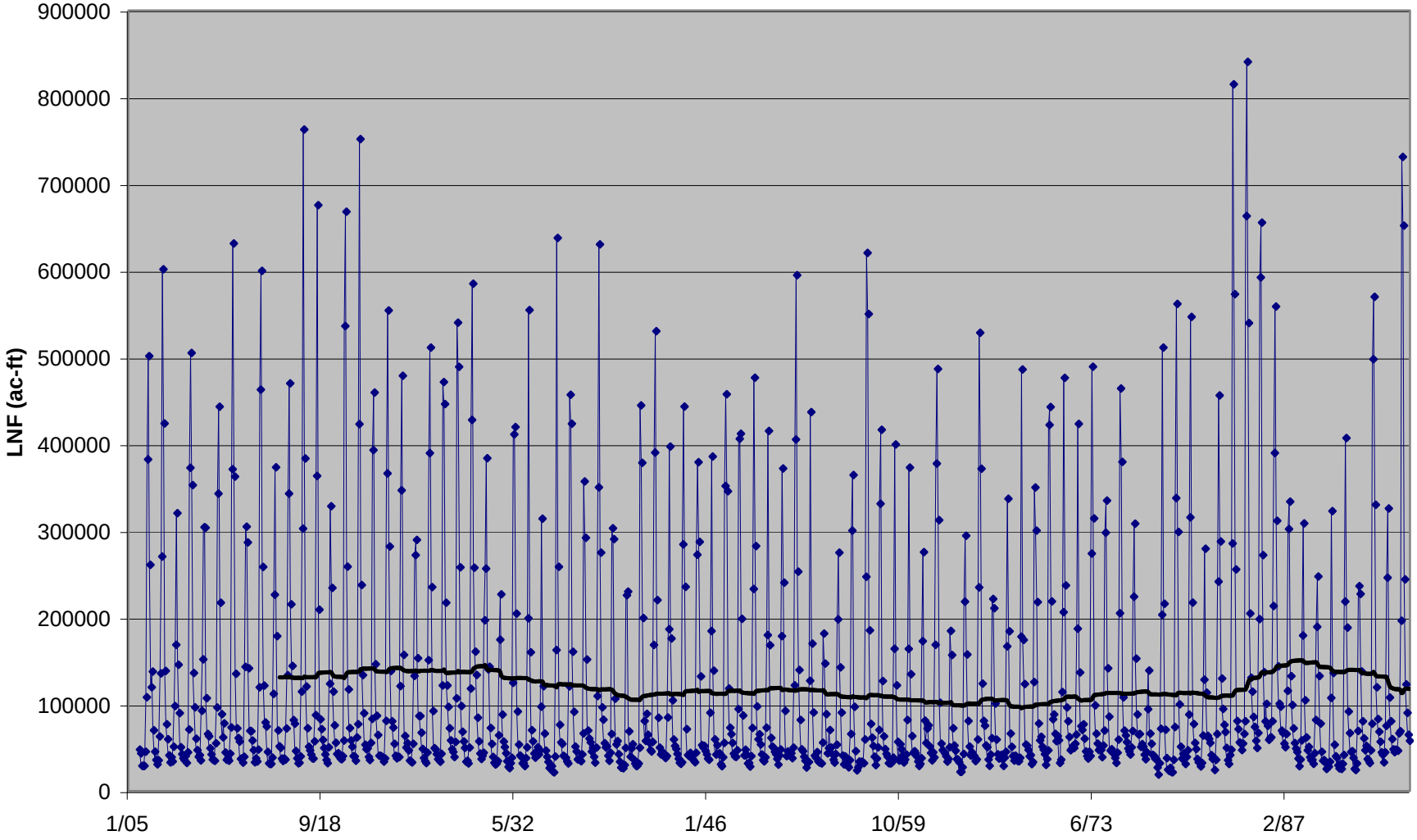


Local Natural Flow Hydrographs

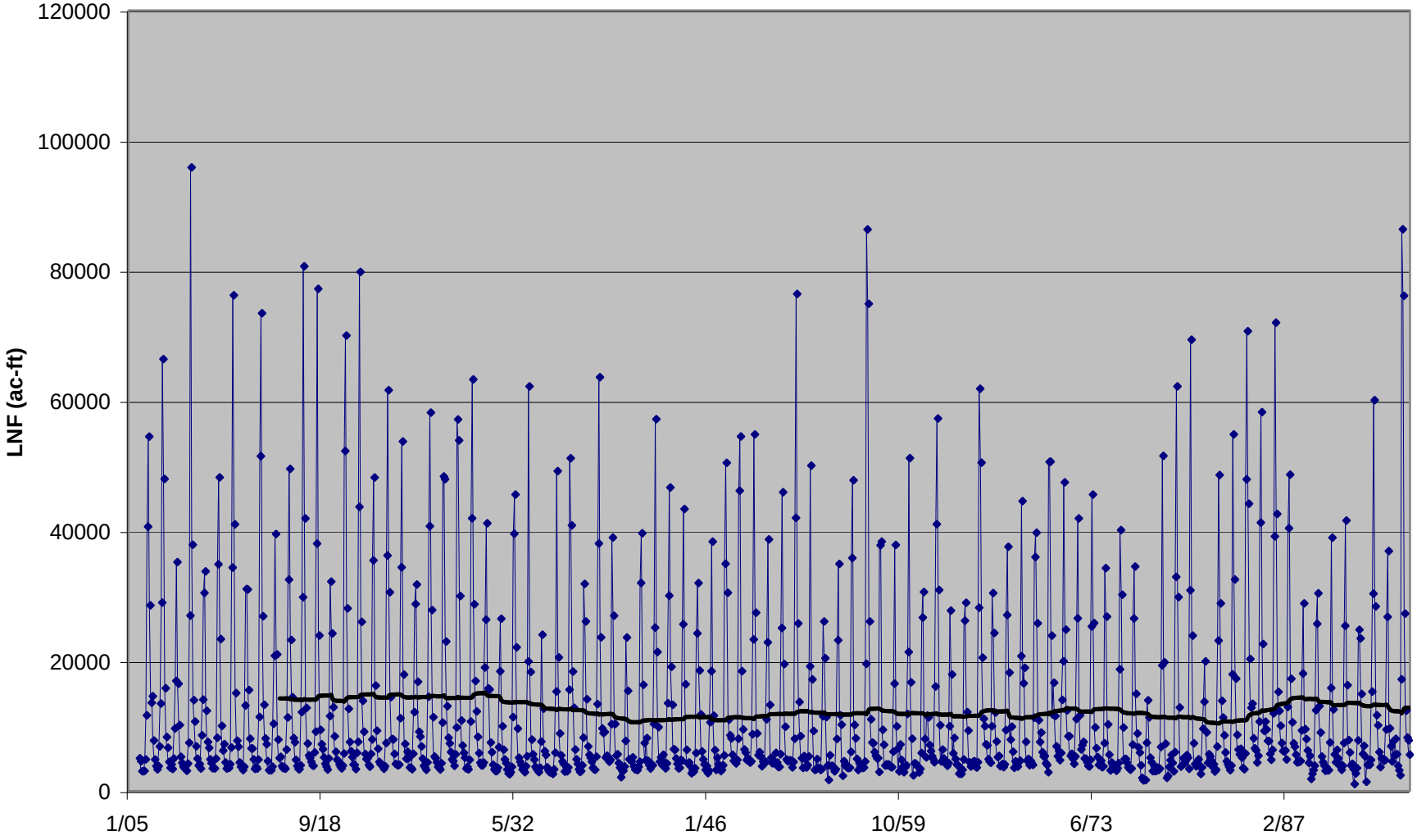
Time Series Data (Local Flows) with 10-yr moving average
Site 1



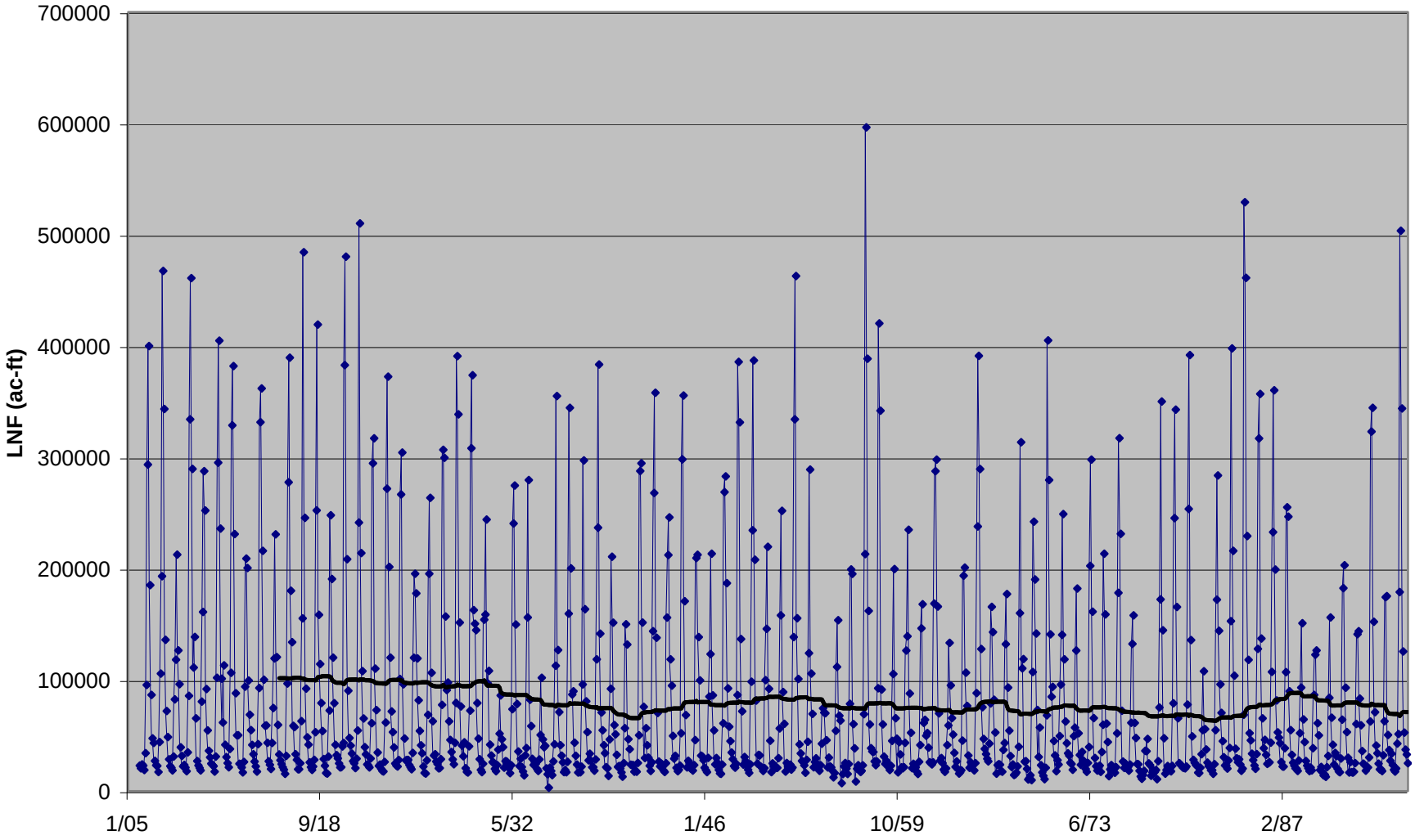
Time Series Data (Local Flows) with 10-yr moving average
Site 2



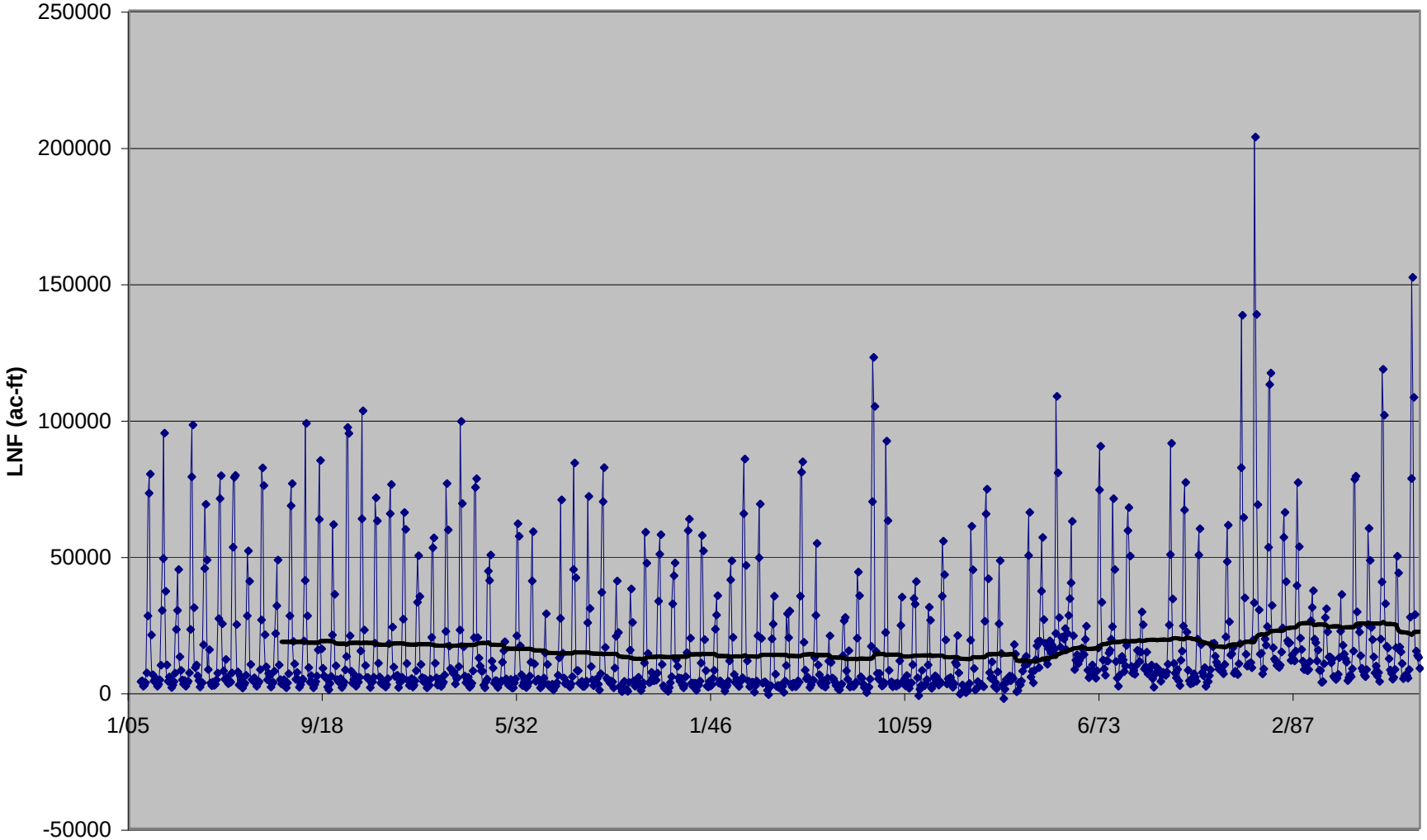
Time Series Data (Local Flows) with 10-yr moving average
Site 3



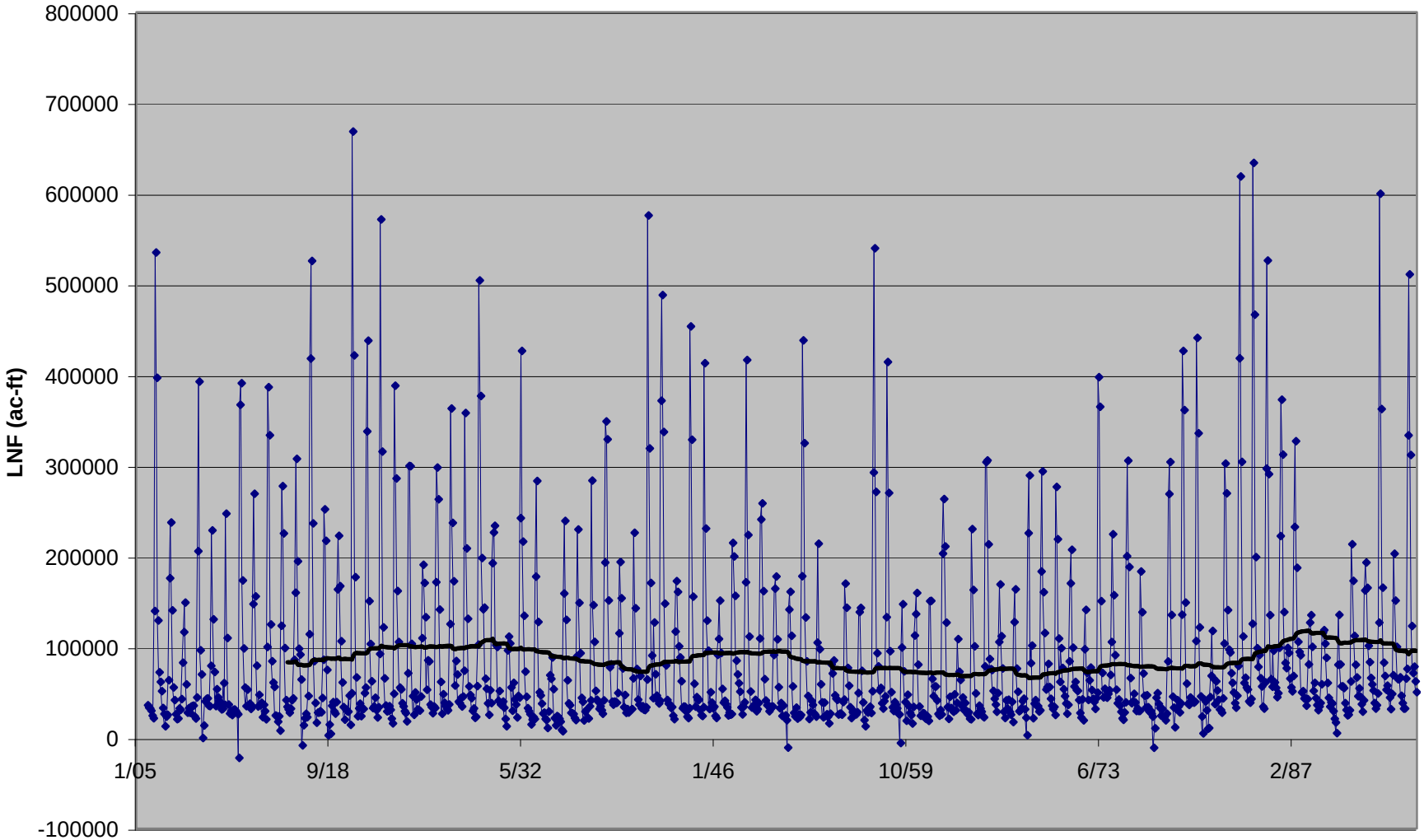
Time Series Data (Local Flows) with 10-yr moving average
Site 4



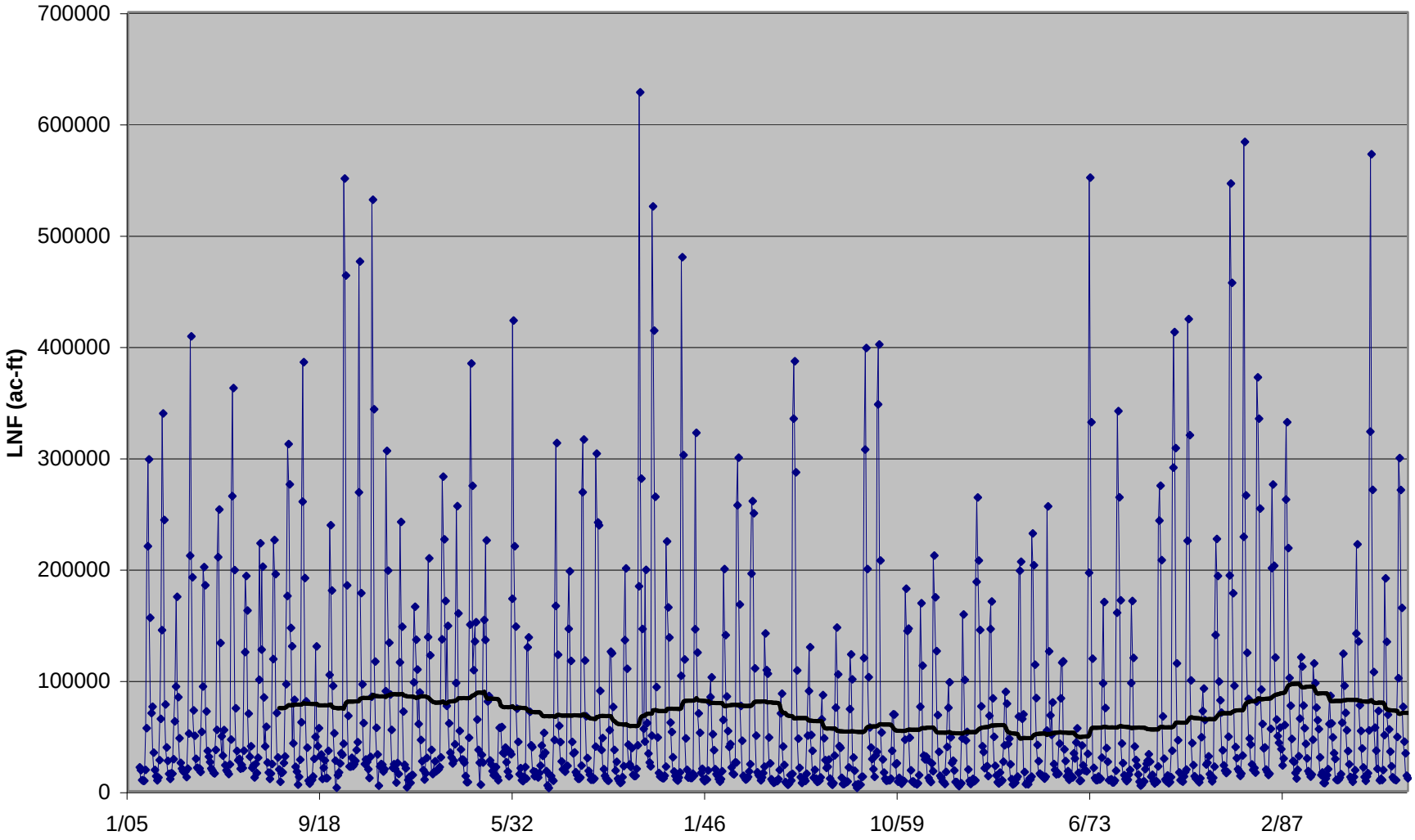
Time Series Data (Local Flows) with 10-yr moving average
Site 5



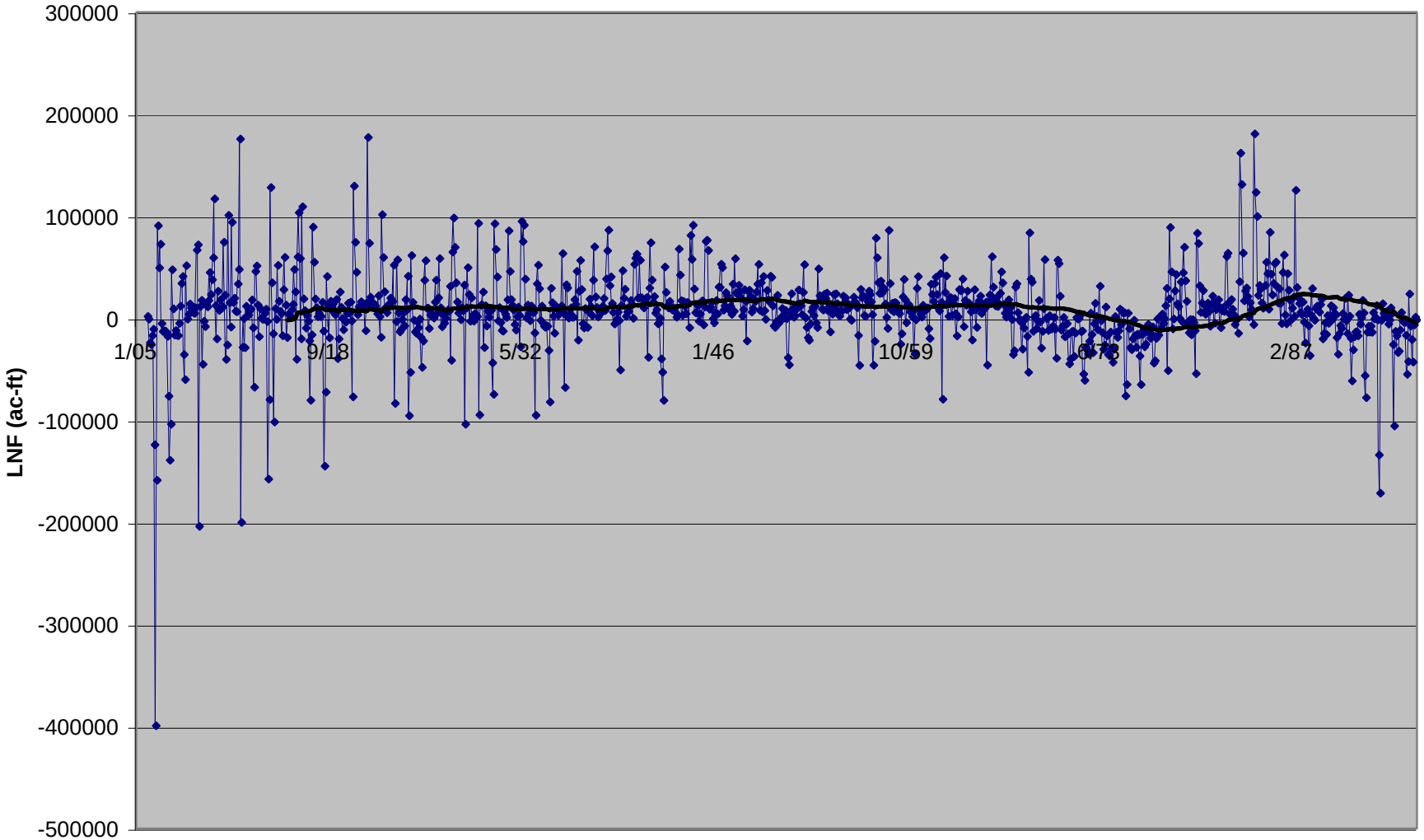
Time Series Data (Local Flows) with 10-yr moving average
Site 6



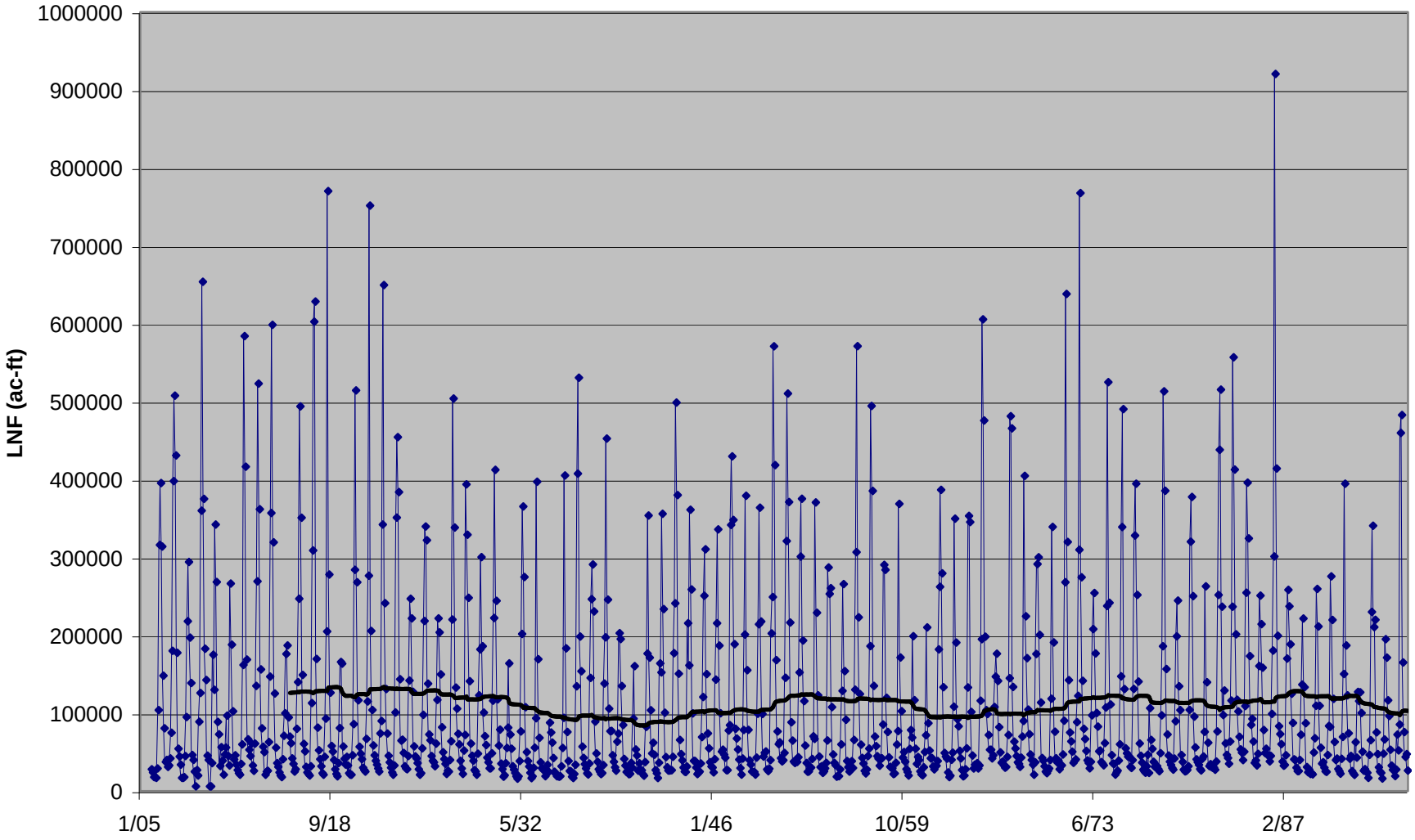
Time Series Data (Local Flows) with 10-yr moving average
Site 7



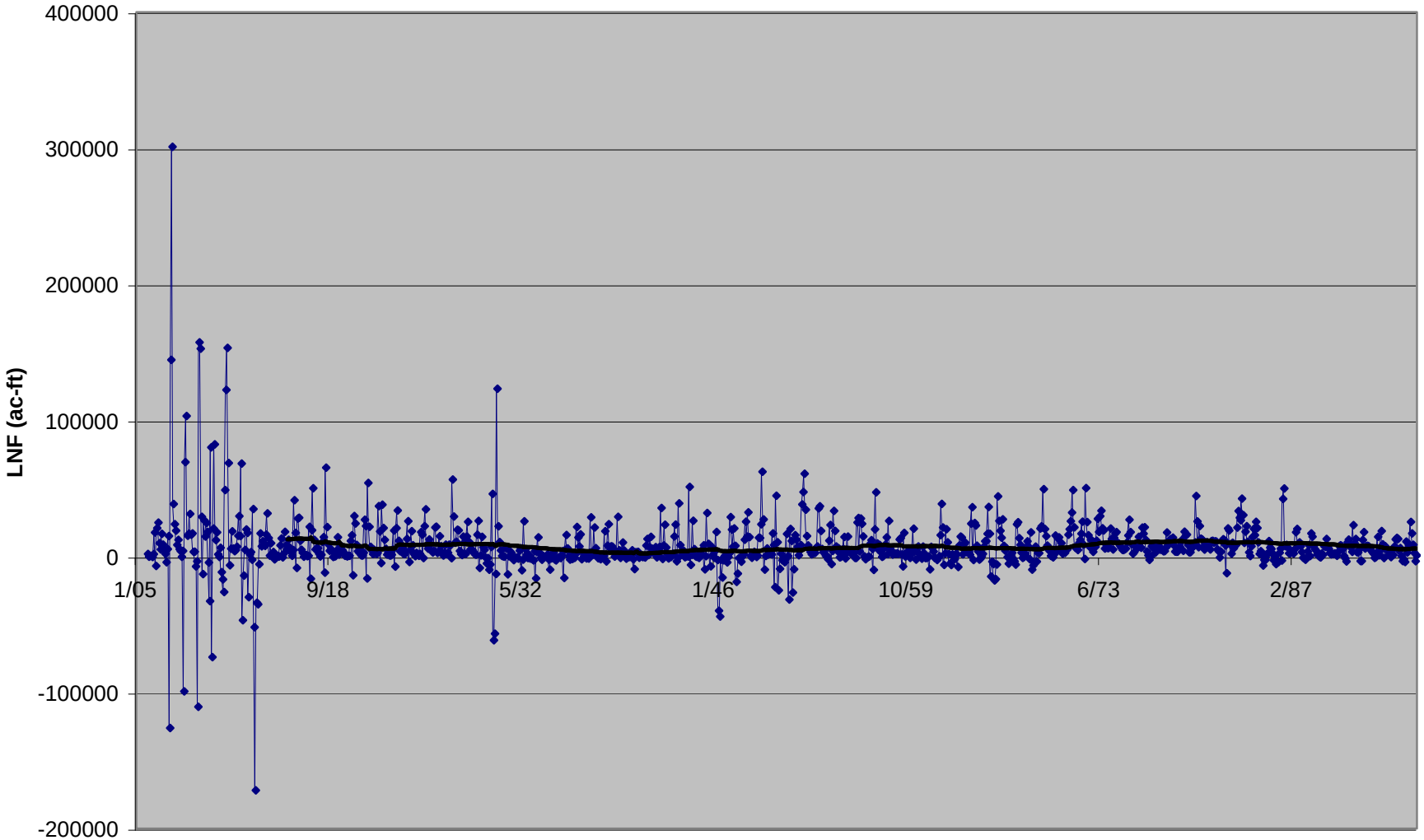
Time Series Data (Local Flows) with 10-yr moving average
Site 8



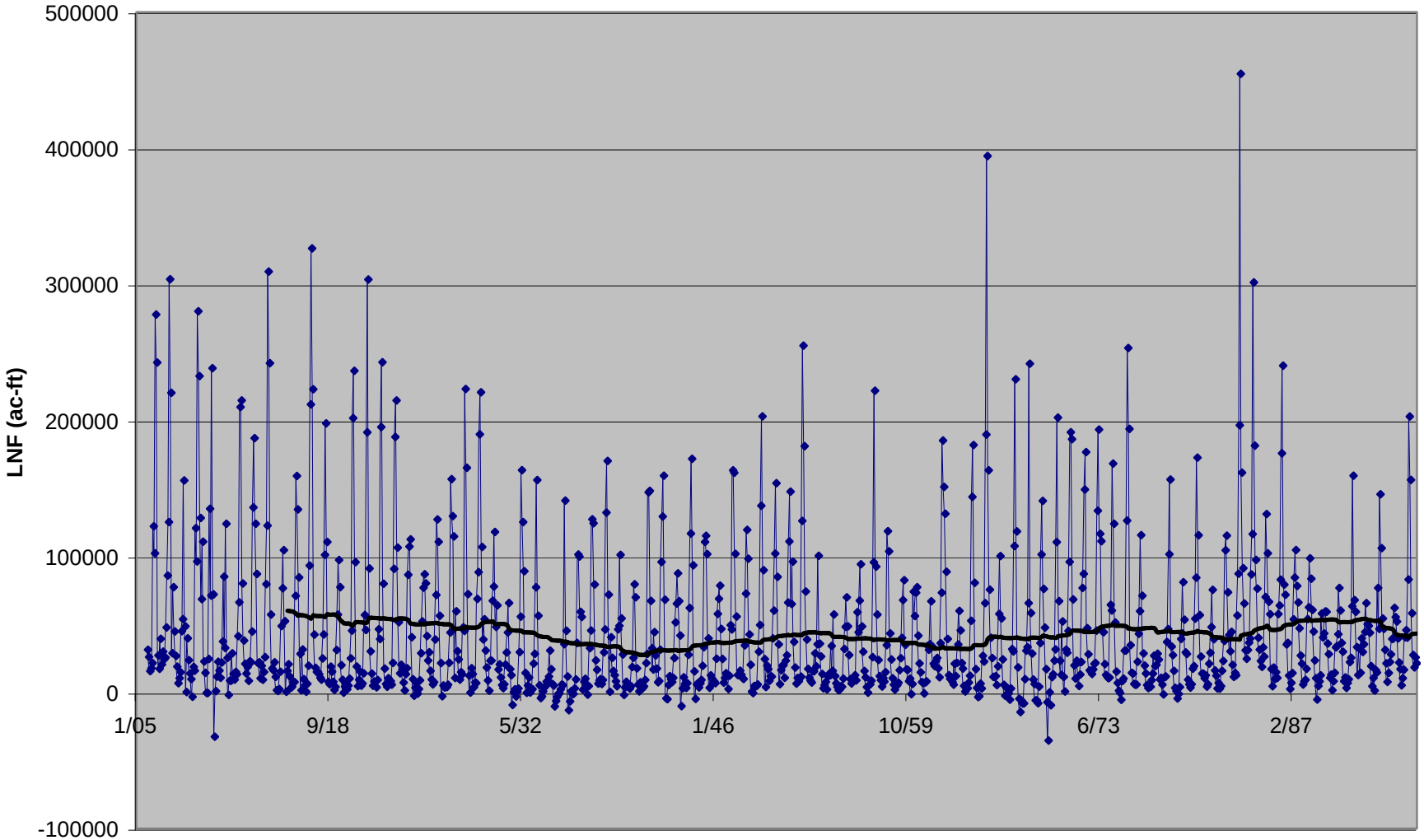
Time Series Data (Local Flows) with 10-yr moving average
Site 9



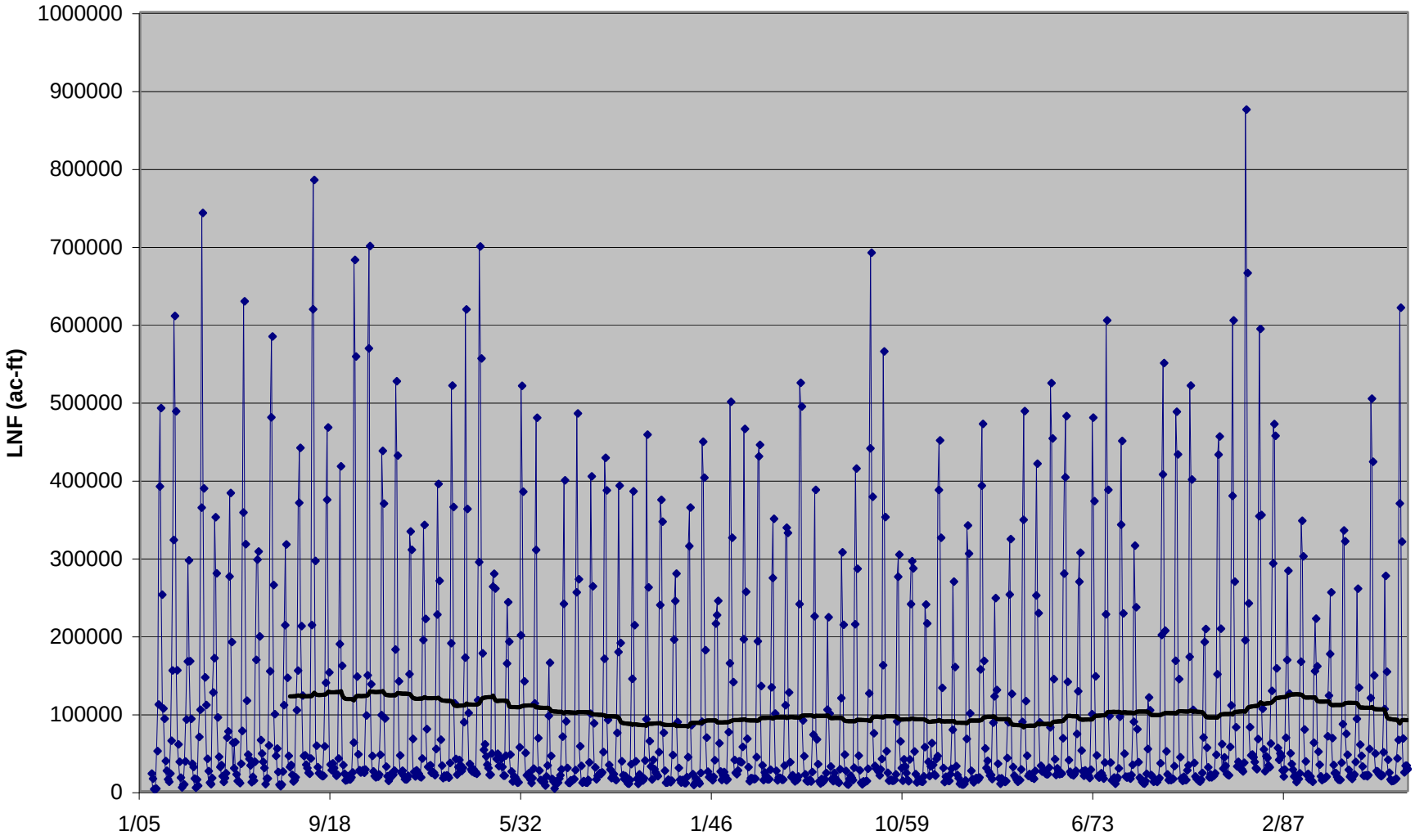
Time Series Data (Local Flows) with 10-yr moving average
Site 10



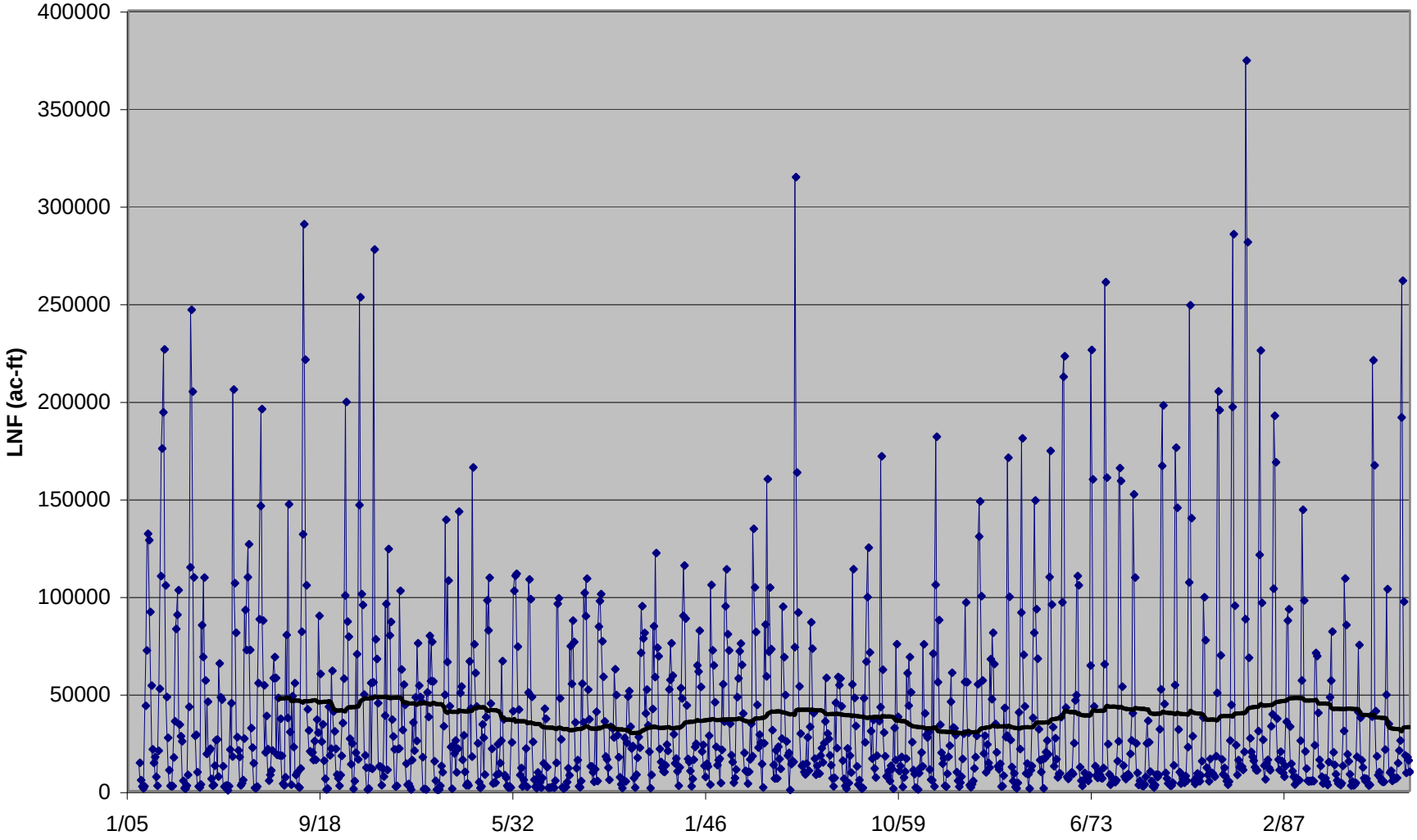
Time Series Data (Local Flows) with 10-yr moving average
Site 11



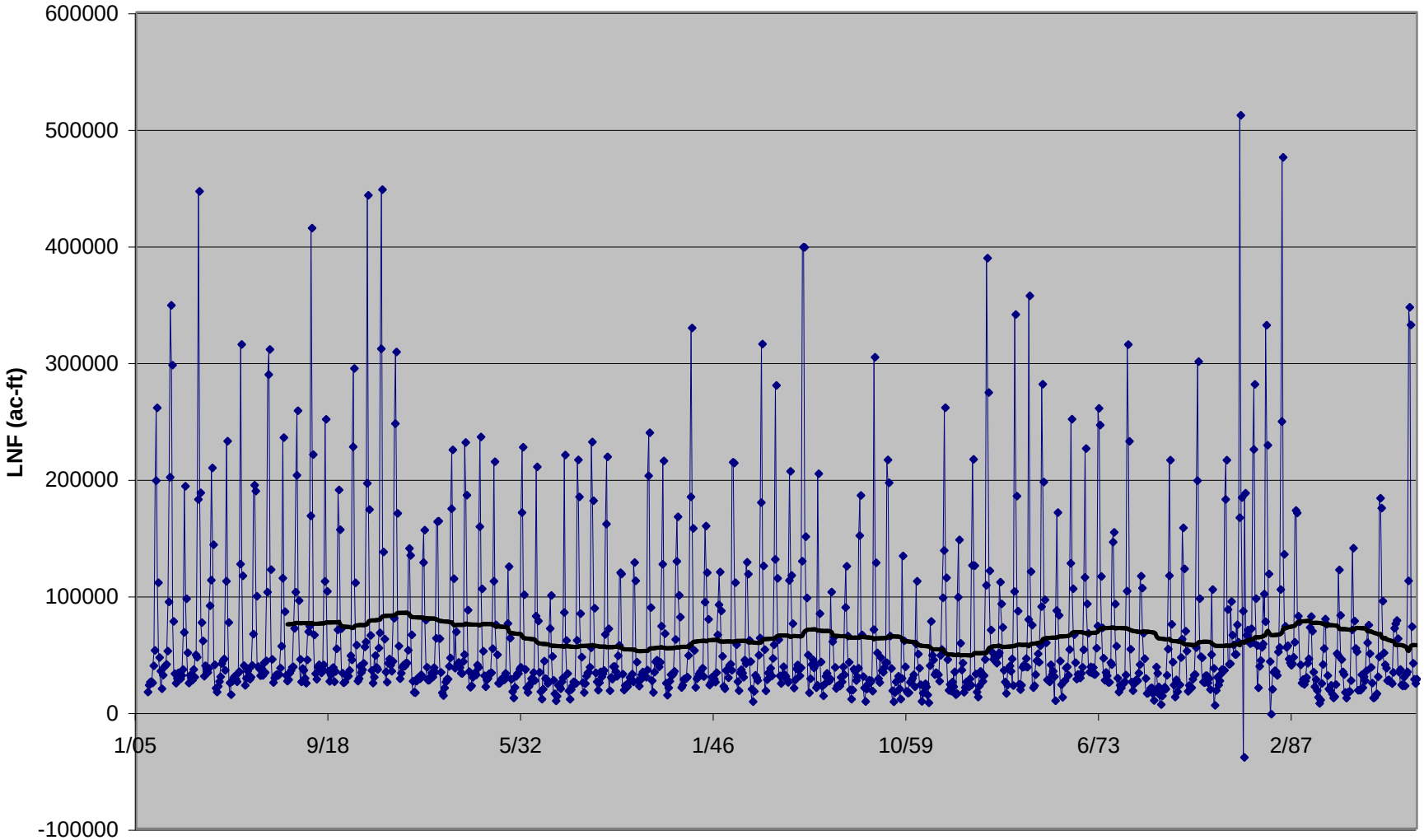
Time Series Data (Local Flows) with 10-yr moving average
Site 12



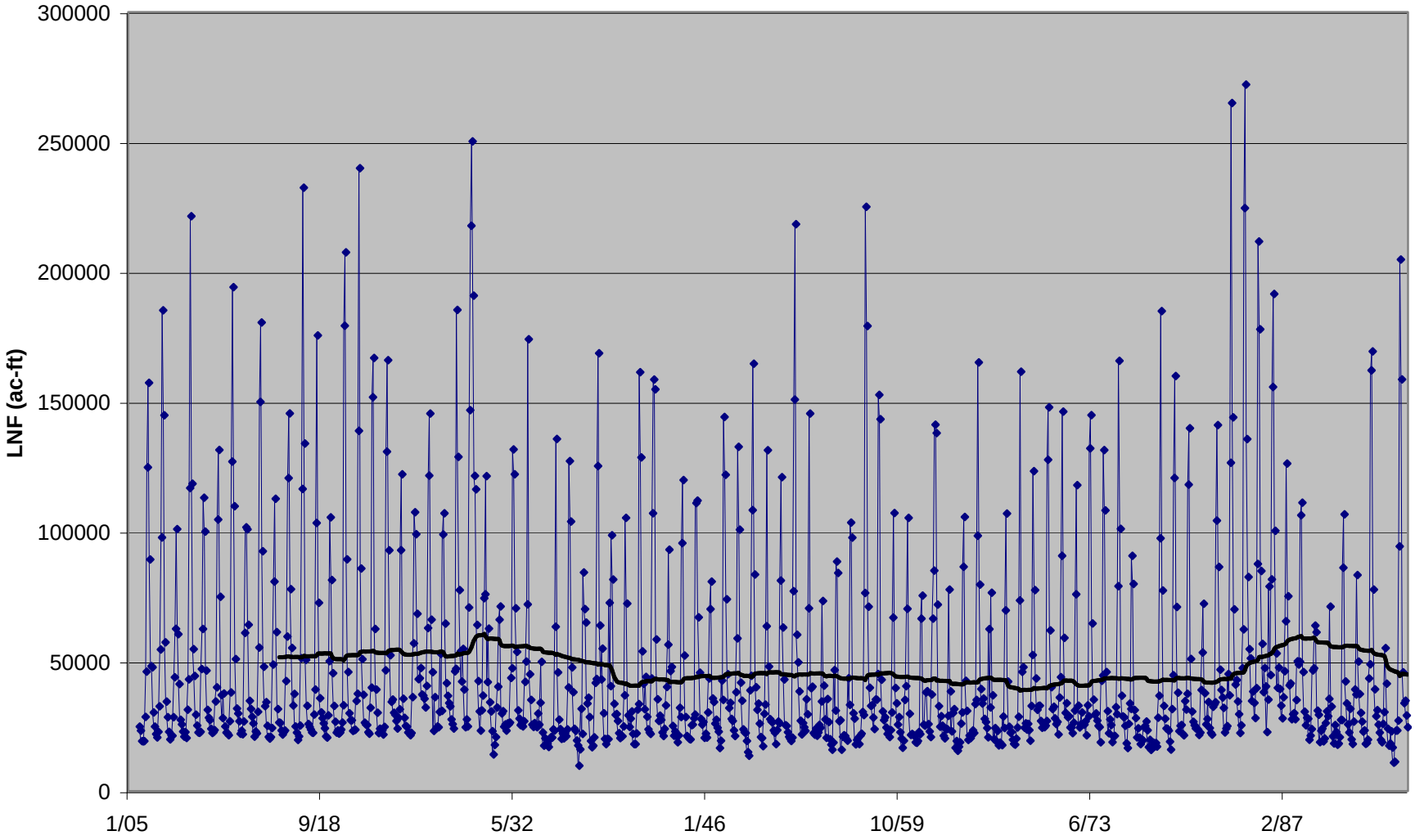
Time Series Data (Local Flows) with 10-yr moving average
Site 13



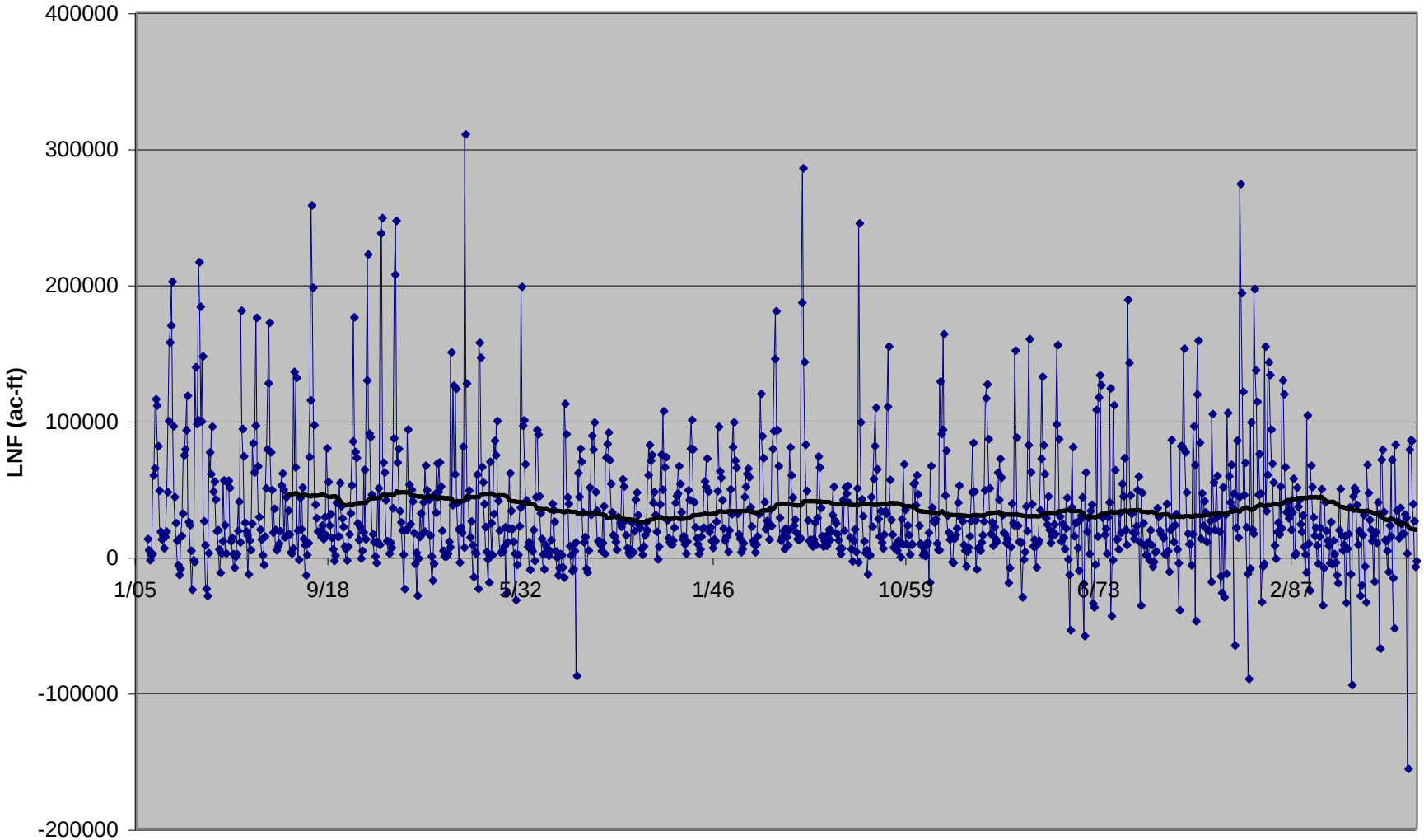
Time Series Data (Local Flows) with 10-yr moving average
Site 14



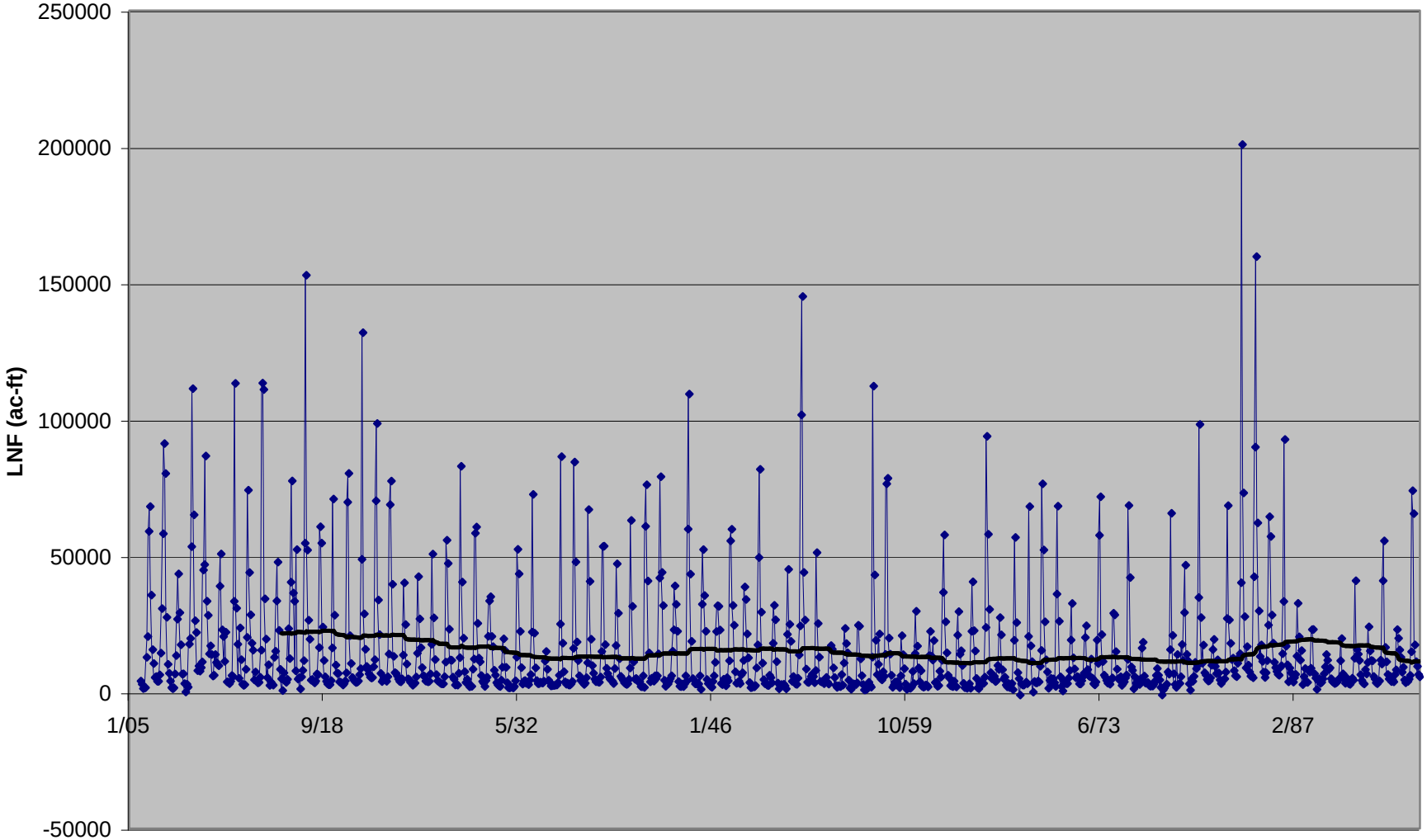
Time Series Data (Local Flows) with 10-yr moving average
Site 15



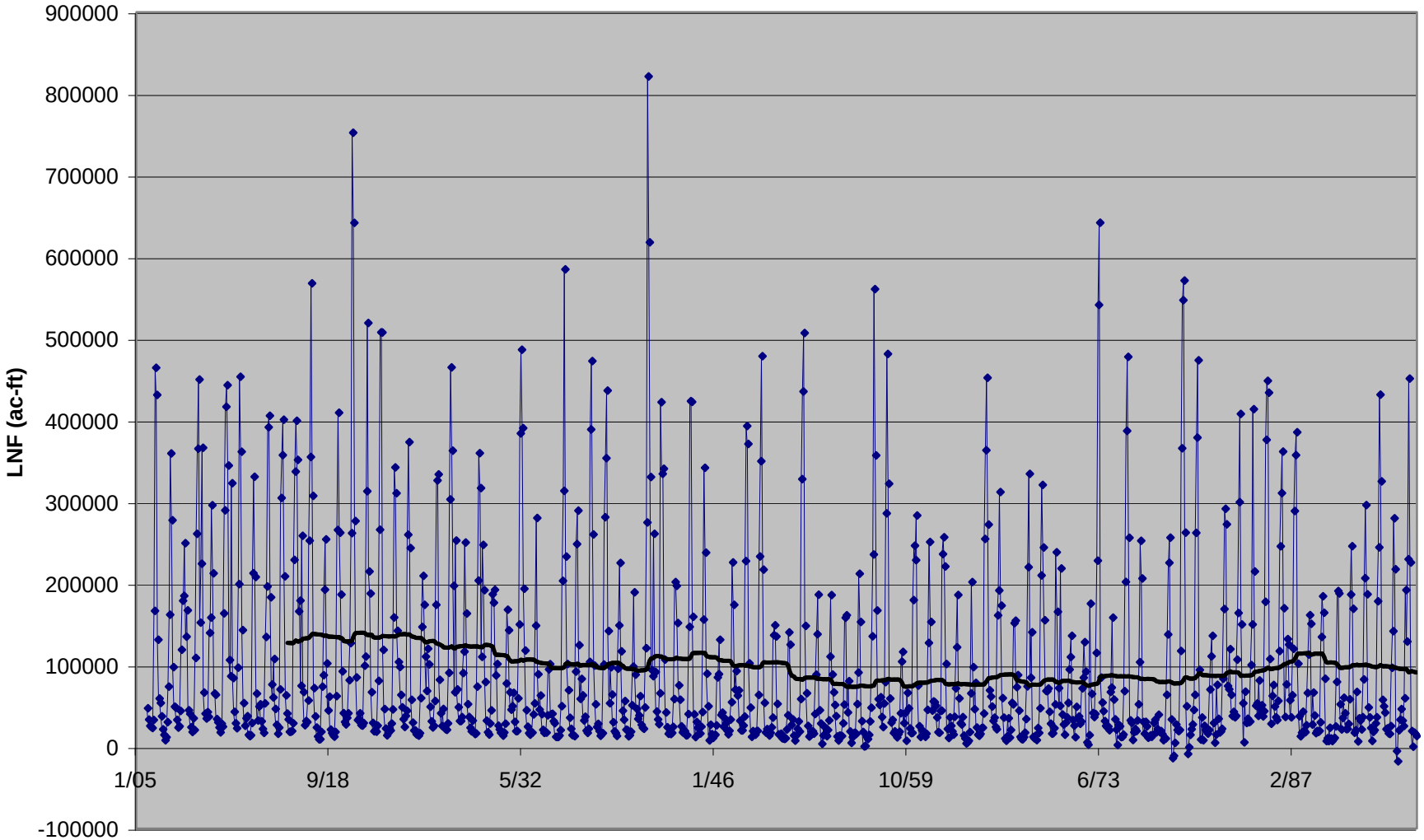
Time Series Data (Local Flows) with 10-yr moving average
Site 16



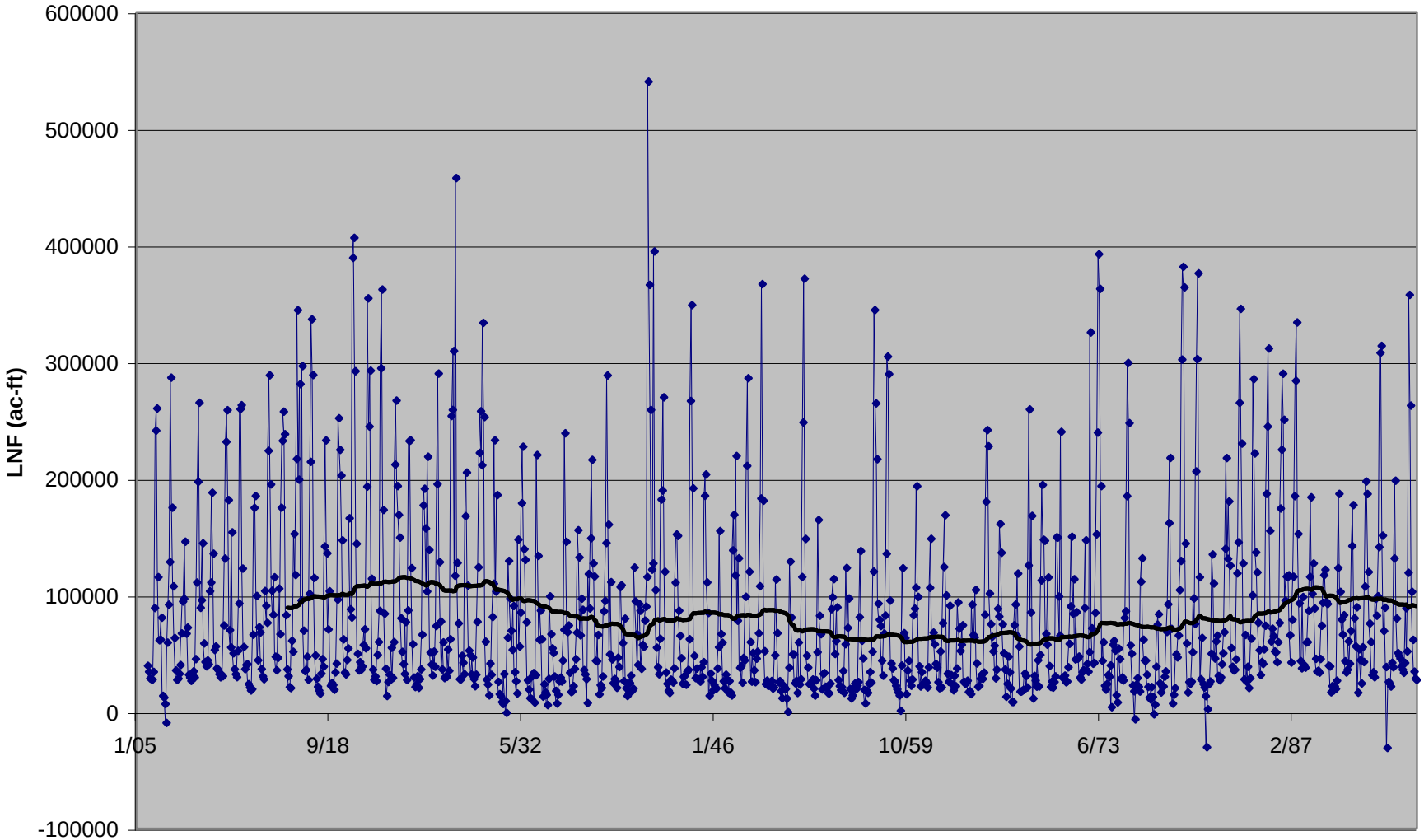
Time Series Data (Local Flows) with 10-yr moving average
Site 17



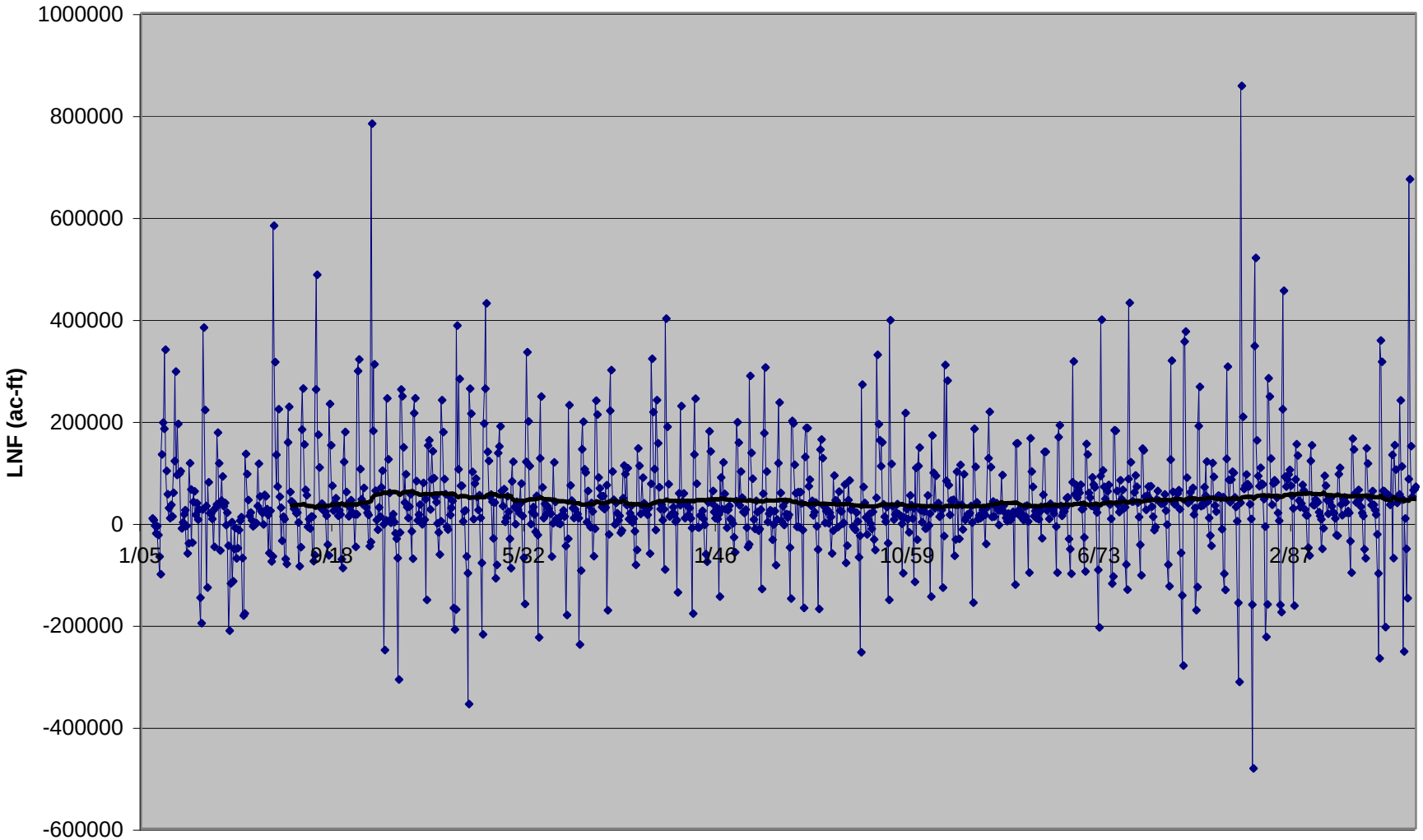
Time Series Data (Local Flows) with 10-yr moving average
Site 18



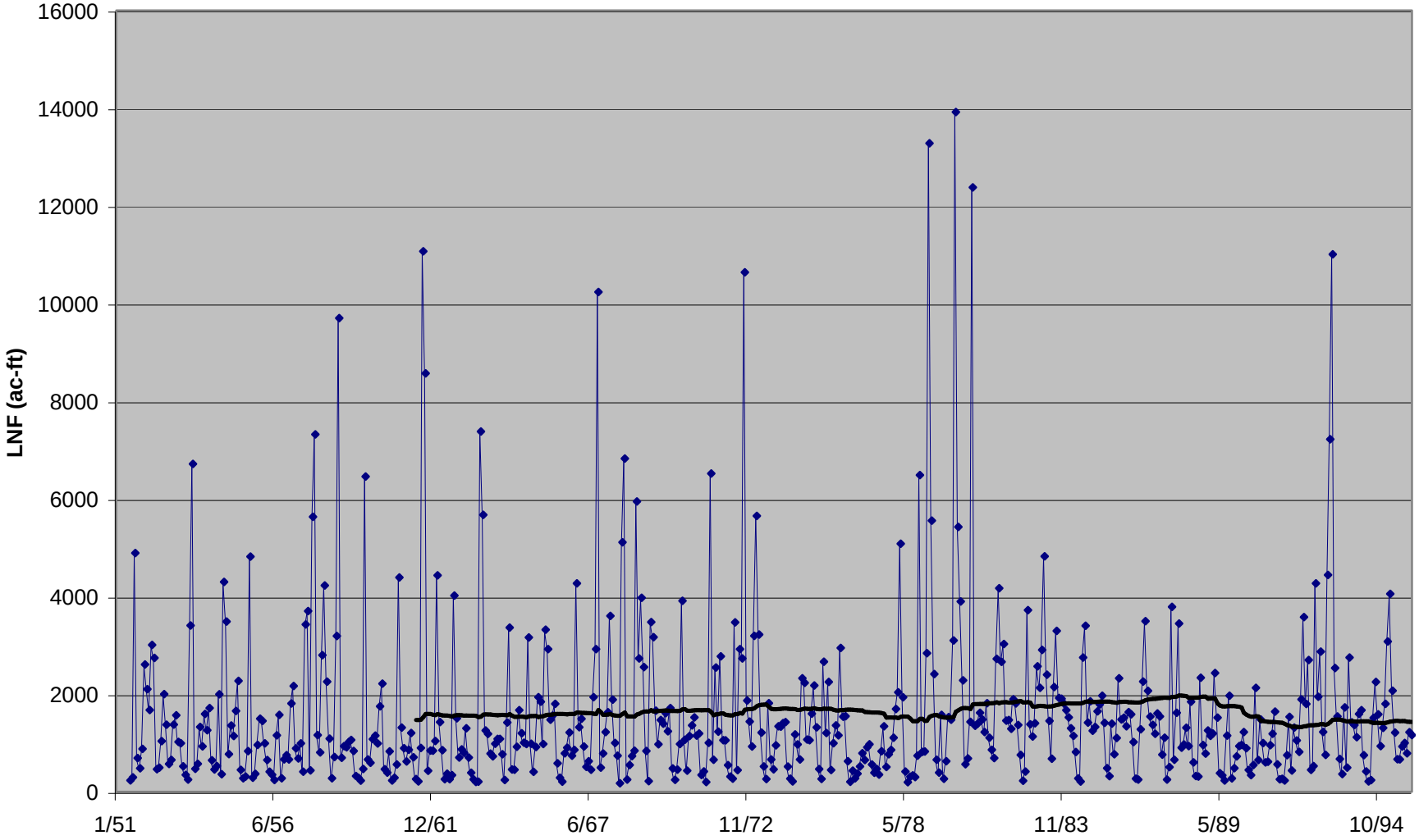
Time Series Data (Local Flows) with 10-yr moving average
Site 19



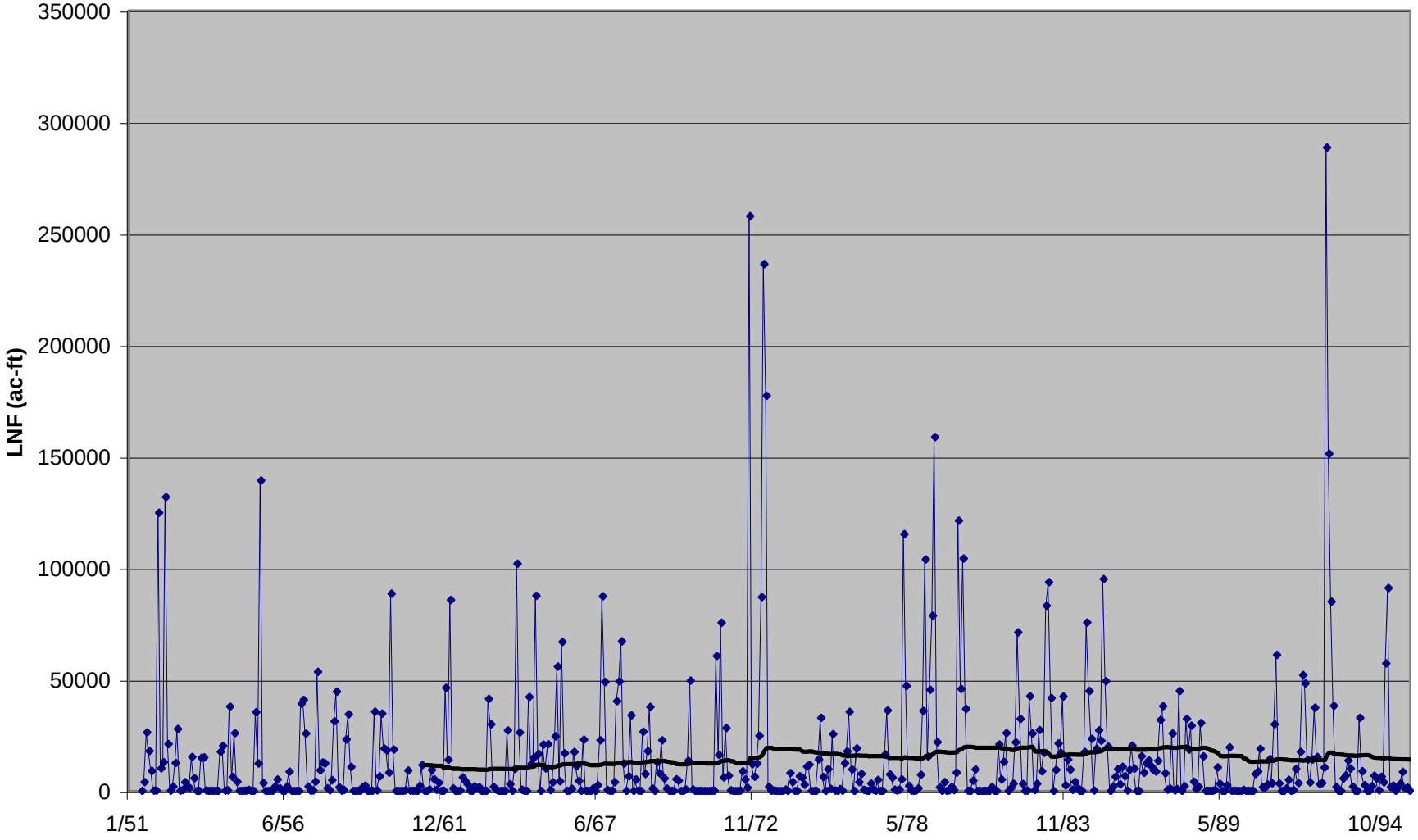
Time Series Data (Local Flows) with 10-yr moving average
Site 20



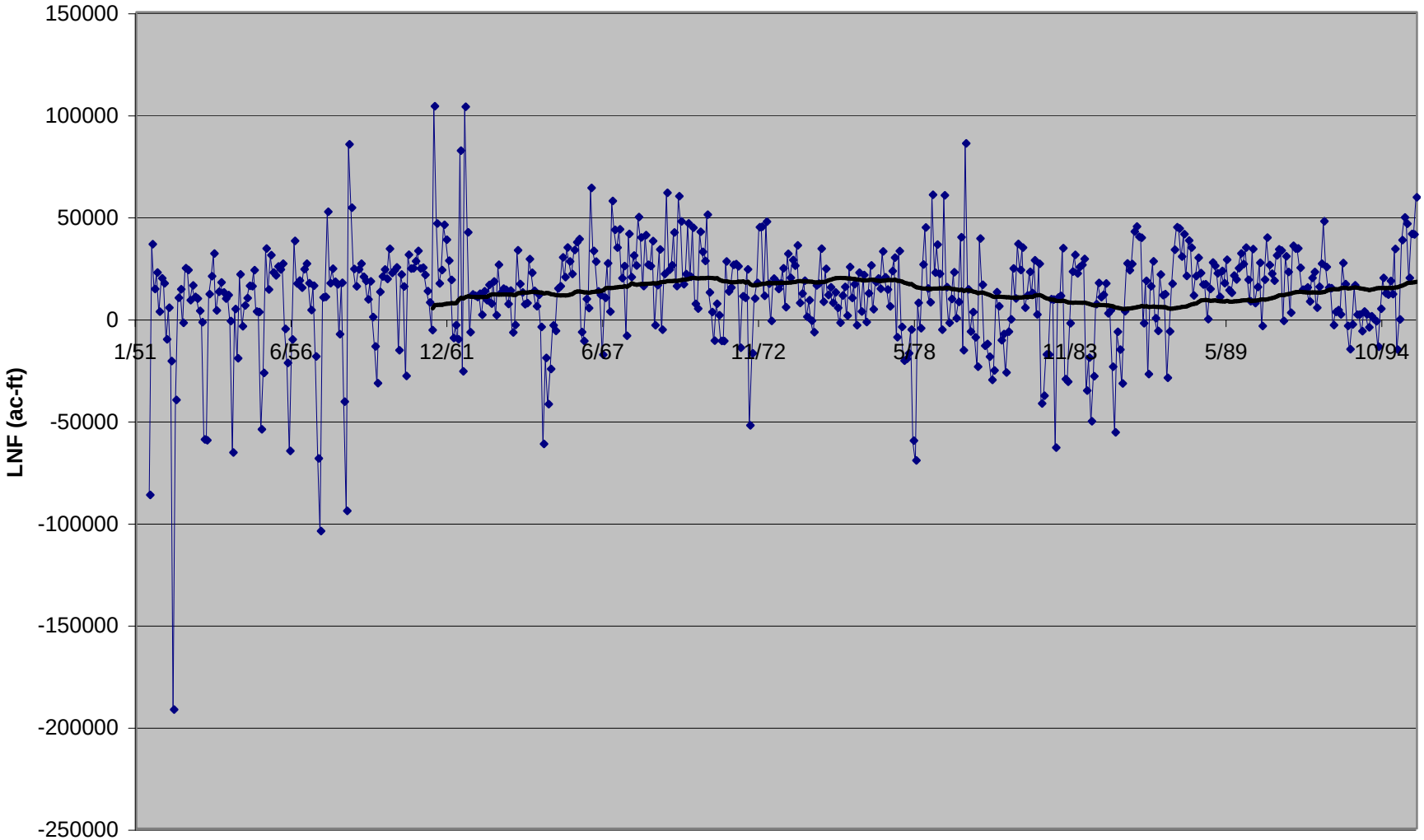
Time Series Data (Local Flows) with 10-yr moving average
Site 21



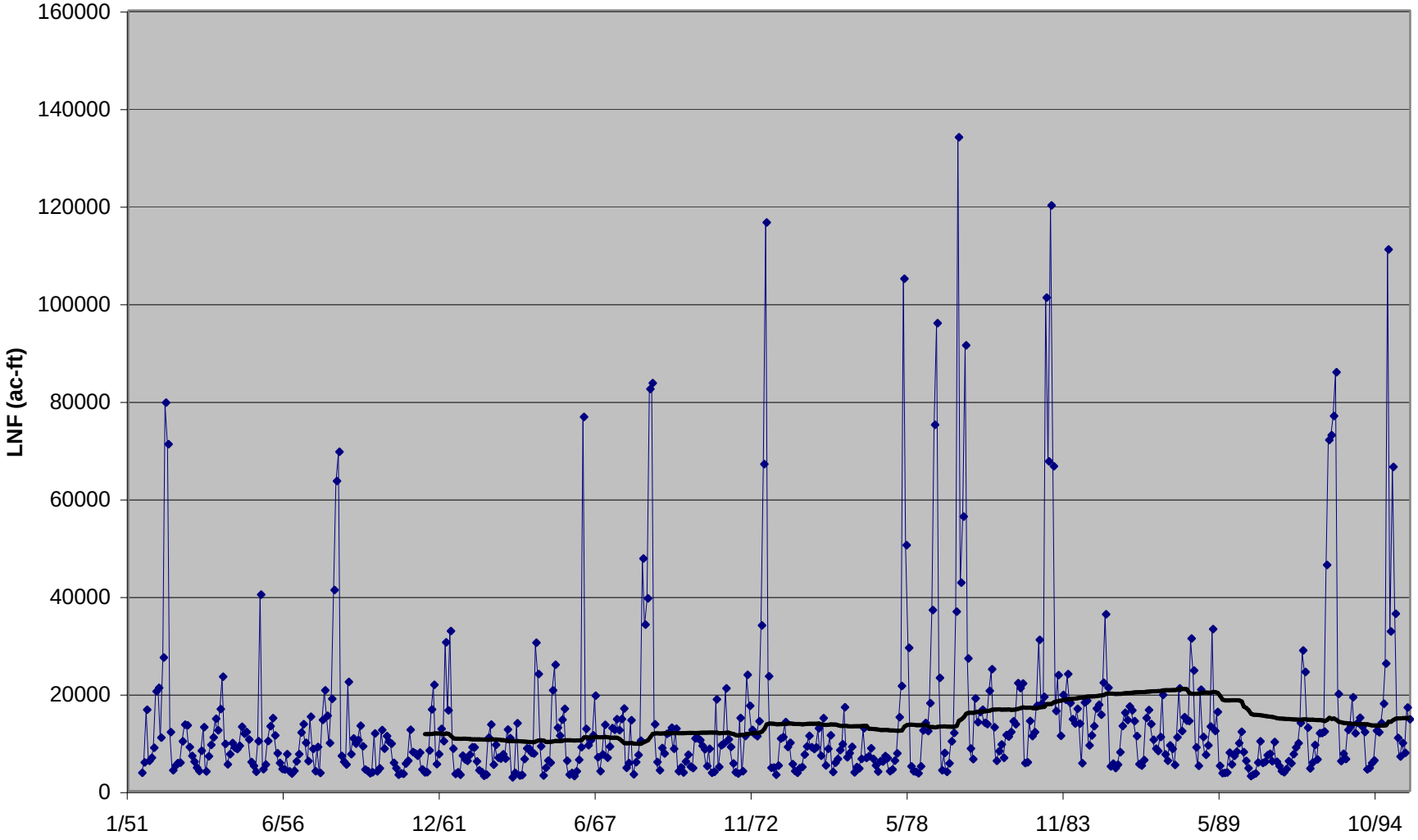
Time Series Data (Local Flows) with 10-yr moving average
Site 22



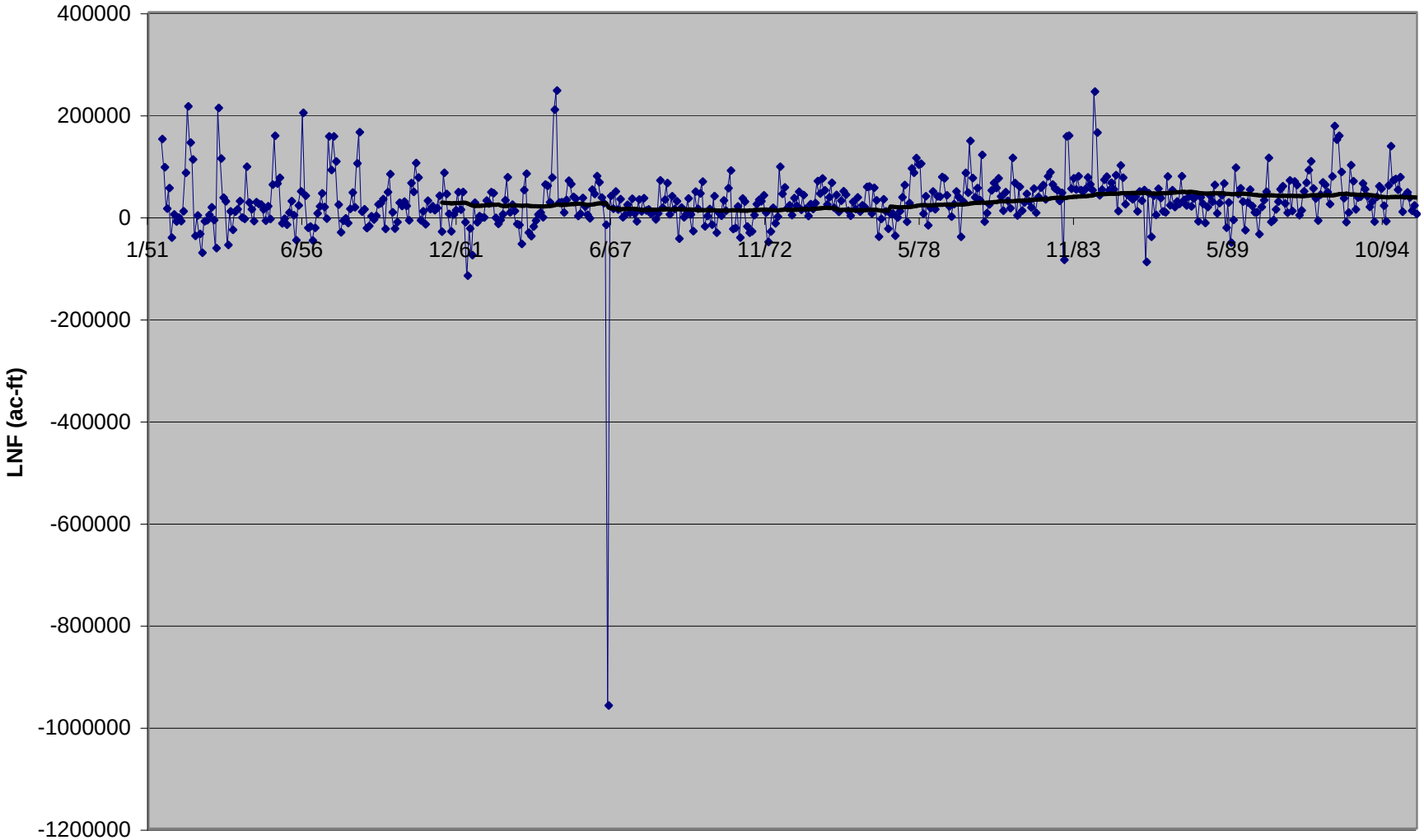
Time Series Data (Local Flows) with 10-yr moving average
Site 23



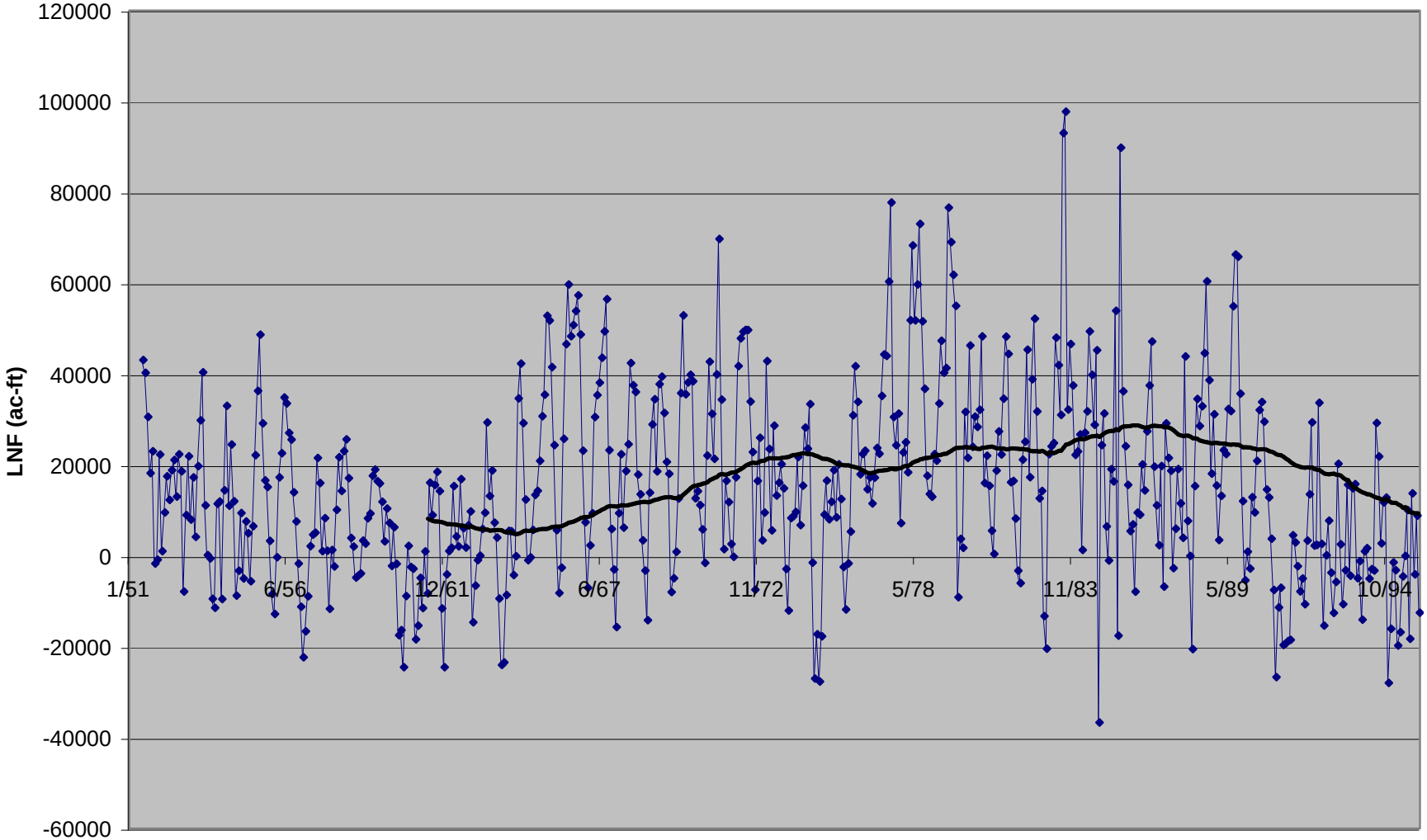
Time Series Data (Local Flows) with 10-yr moving average
Site 24



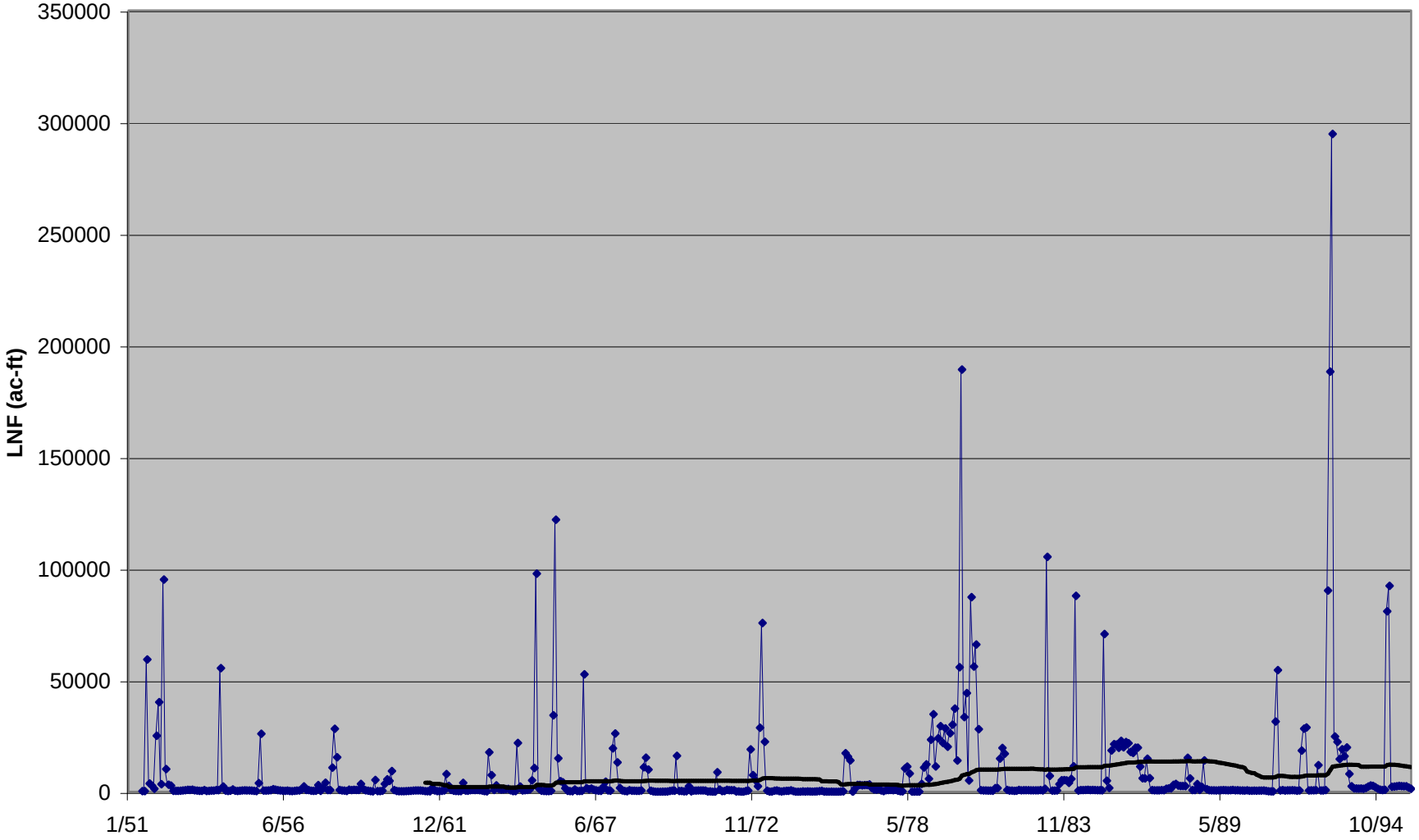
Time Series Data (Local Flows) with 10-yr moving average
Site 25



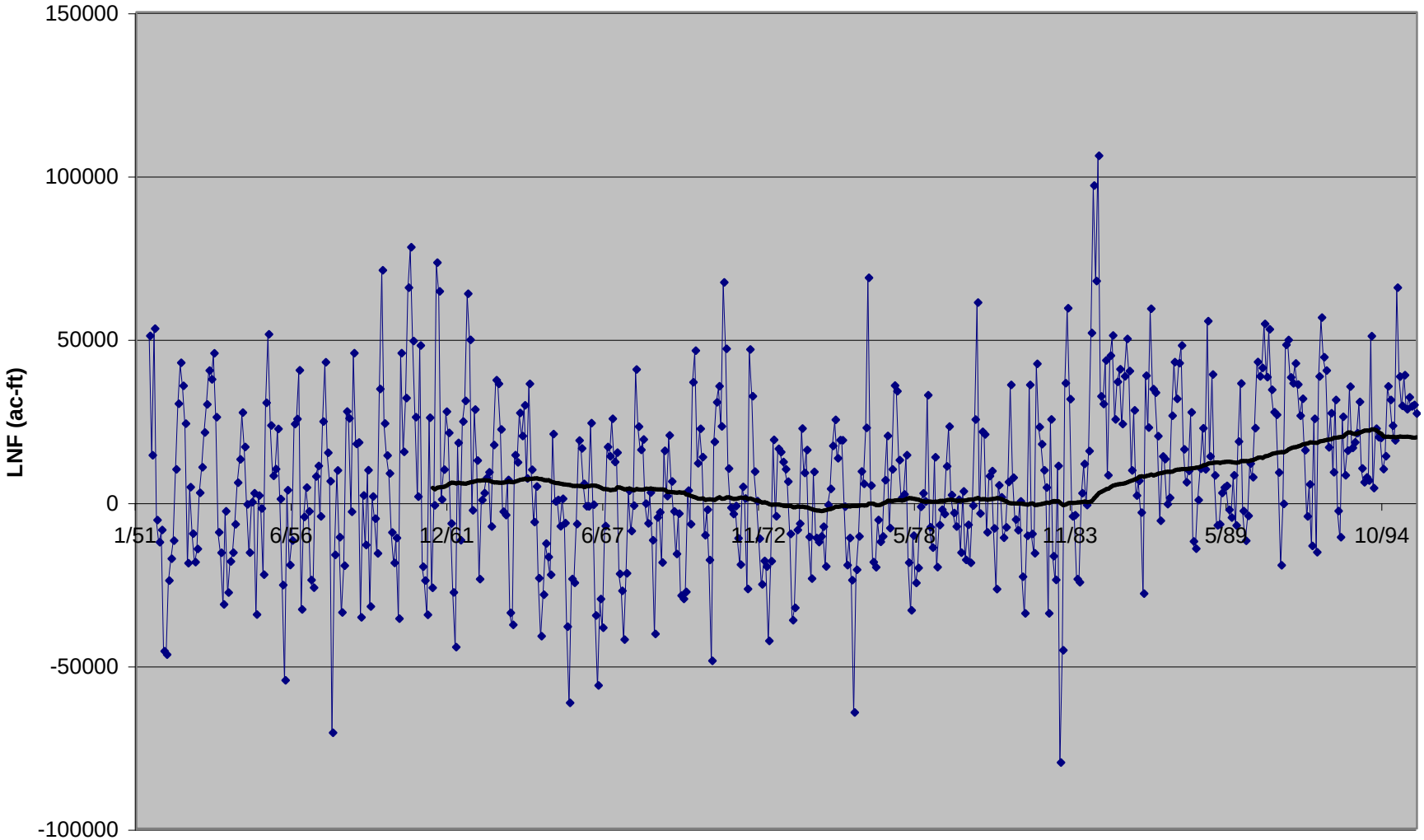
Time Series Data (Local Flows) with 10-yr moving average
Site 26



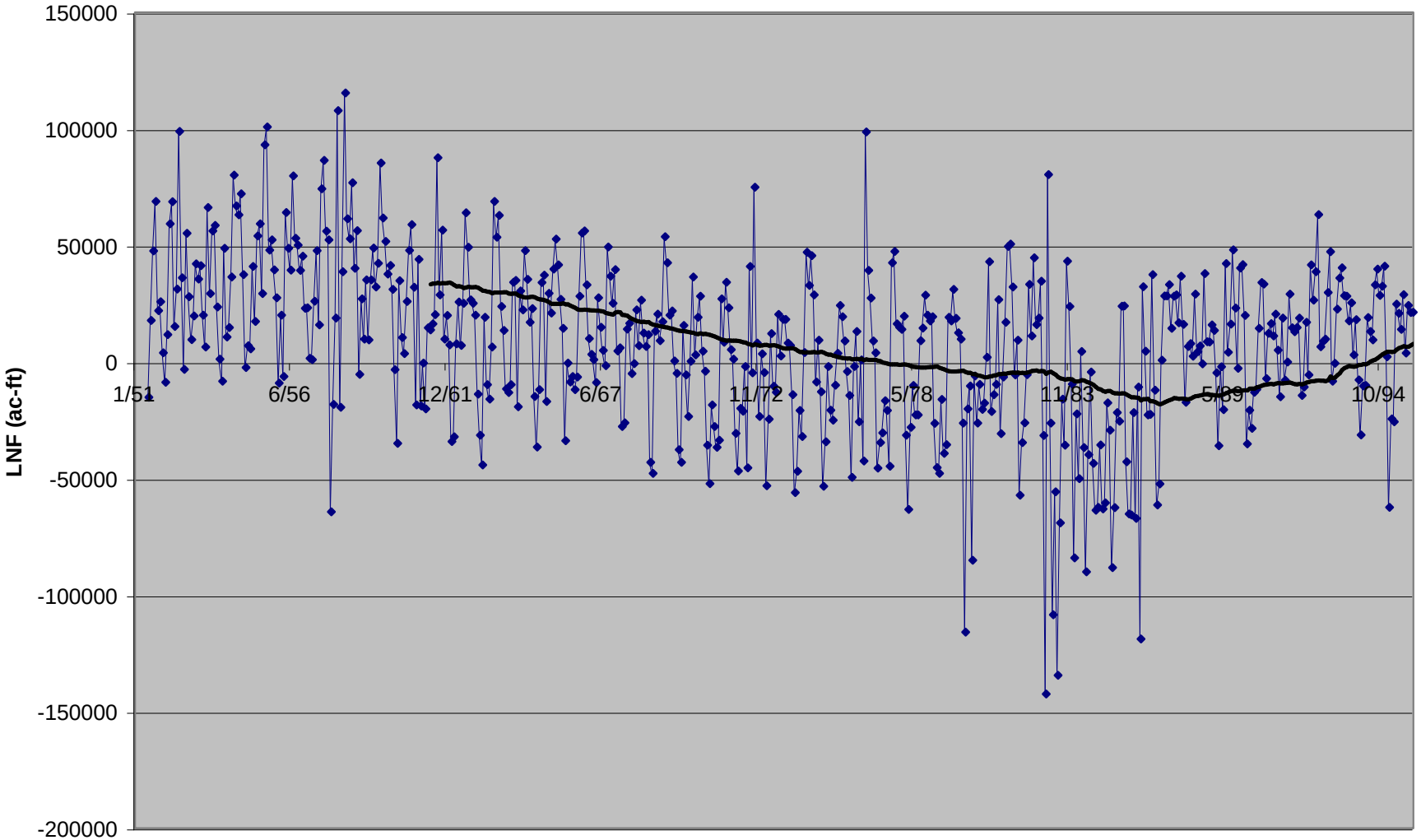
Time Series Data (Local Flows) with 10-yr moving average
Site 27



Time Series Data (Local Flows) with 10-yr moving average
Site 28

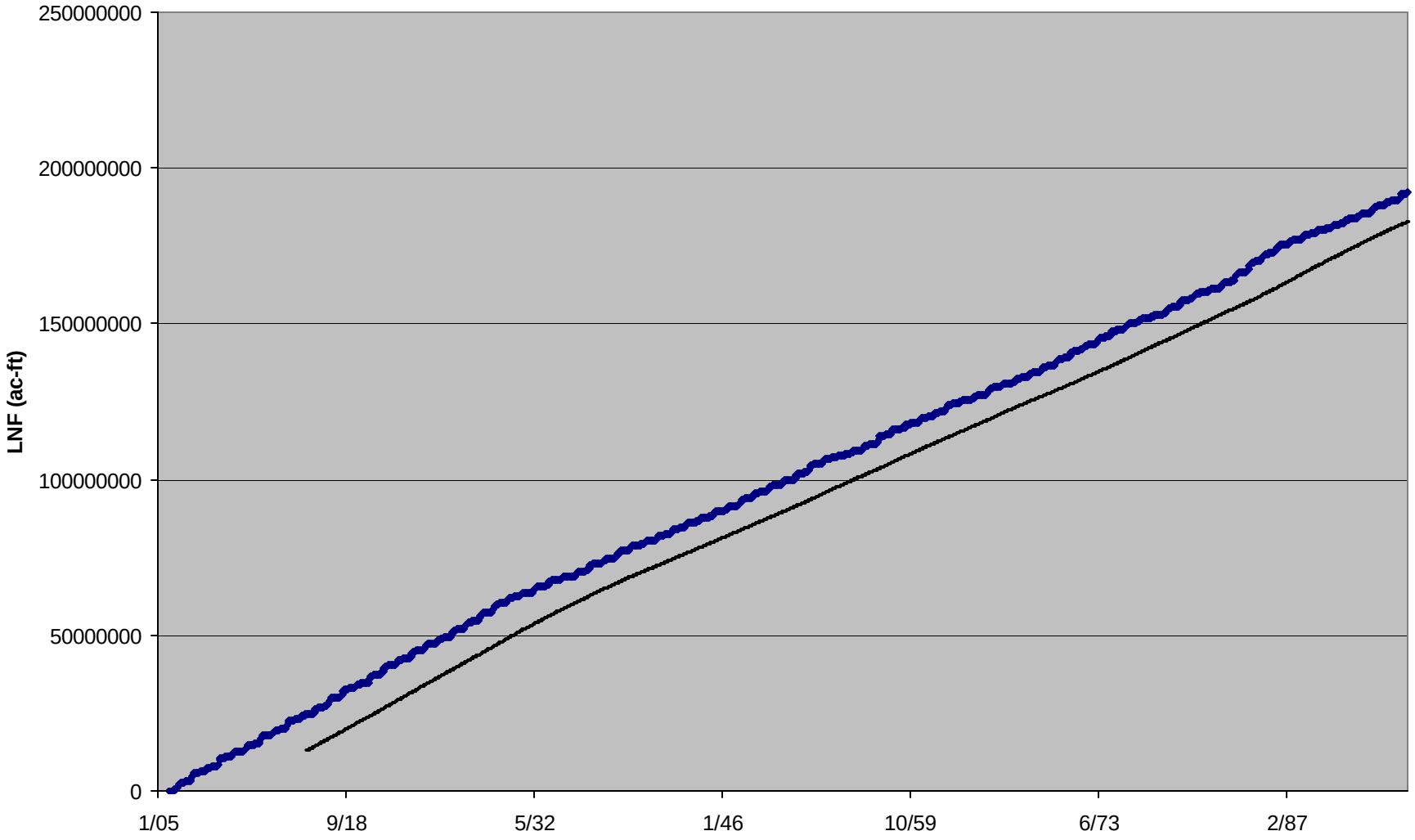


Time Series Data (Local Flows) with 10-yr moving average
Site 29

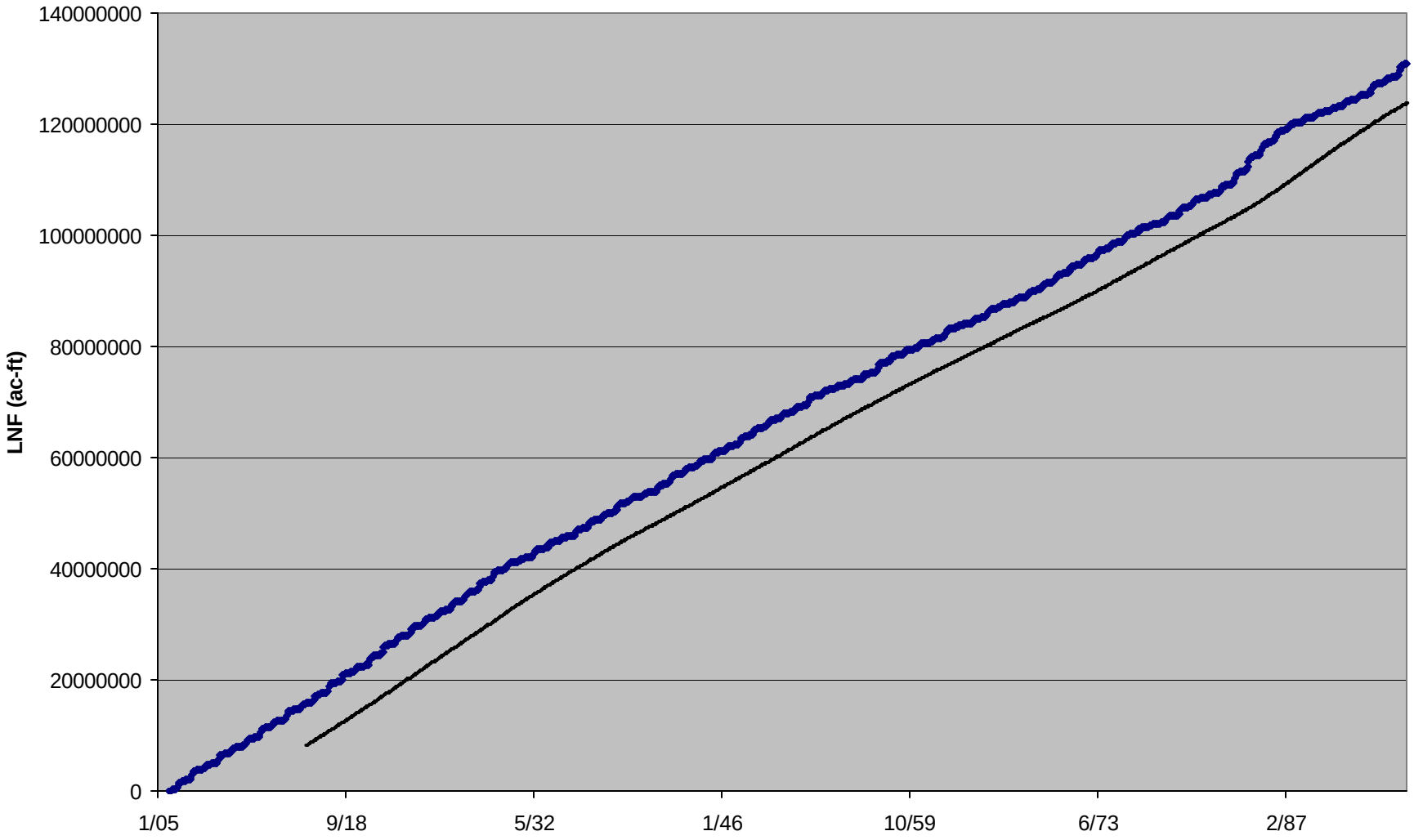


Cumulative Local Natural Flow Plots

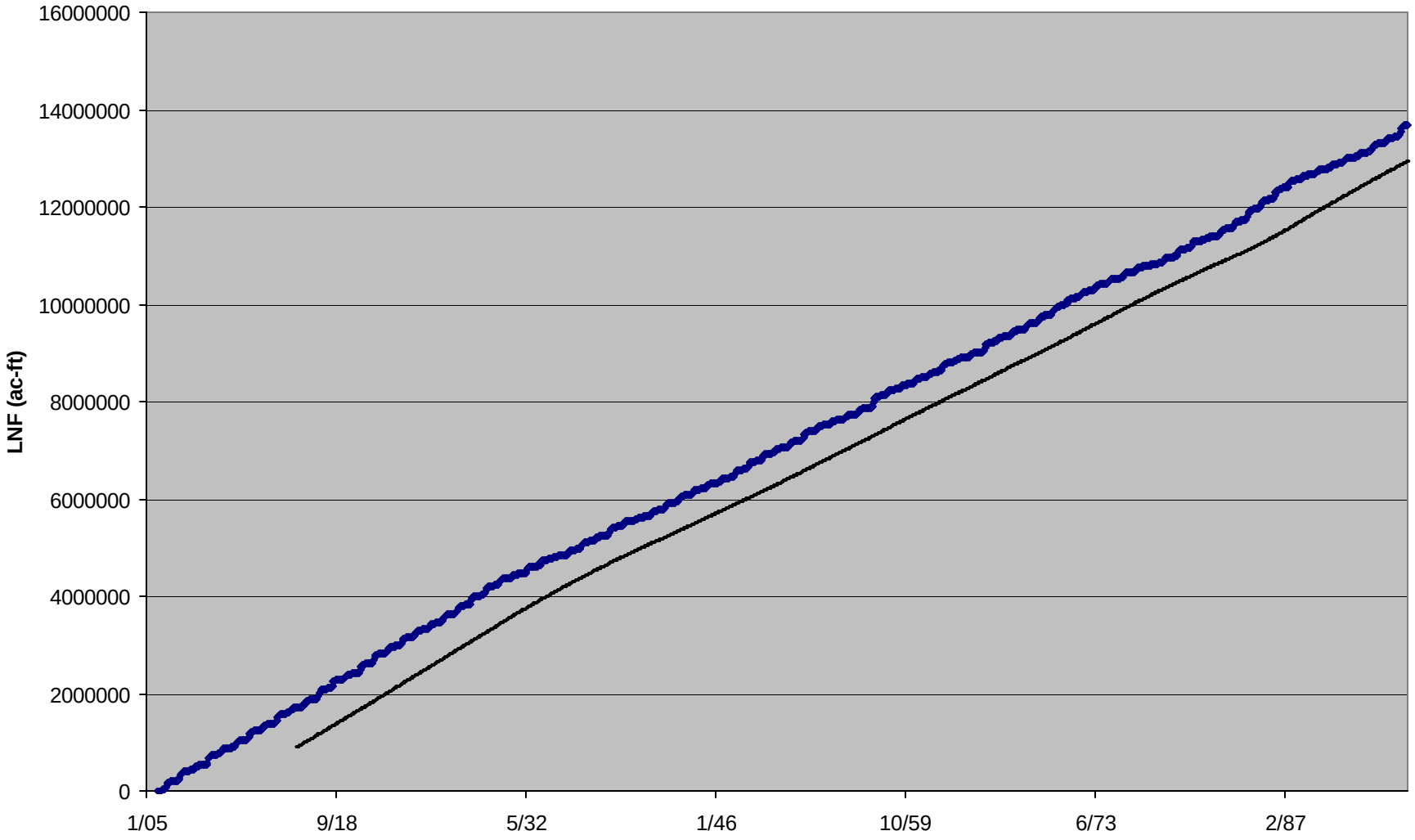
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 1



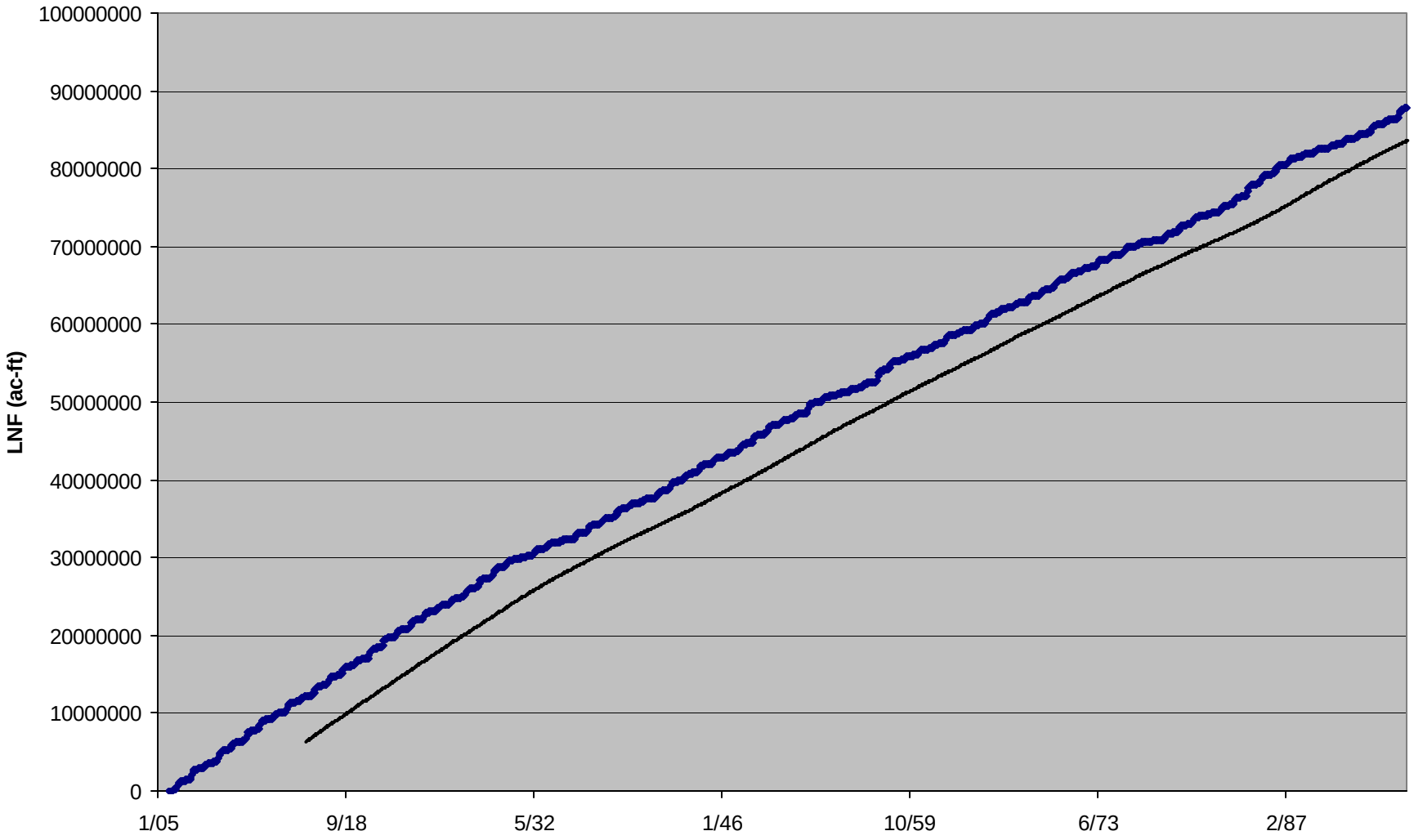
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 2



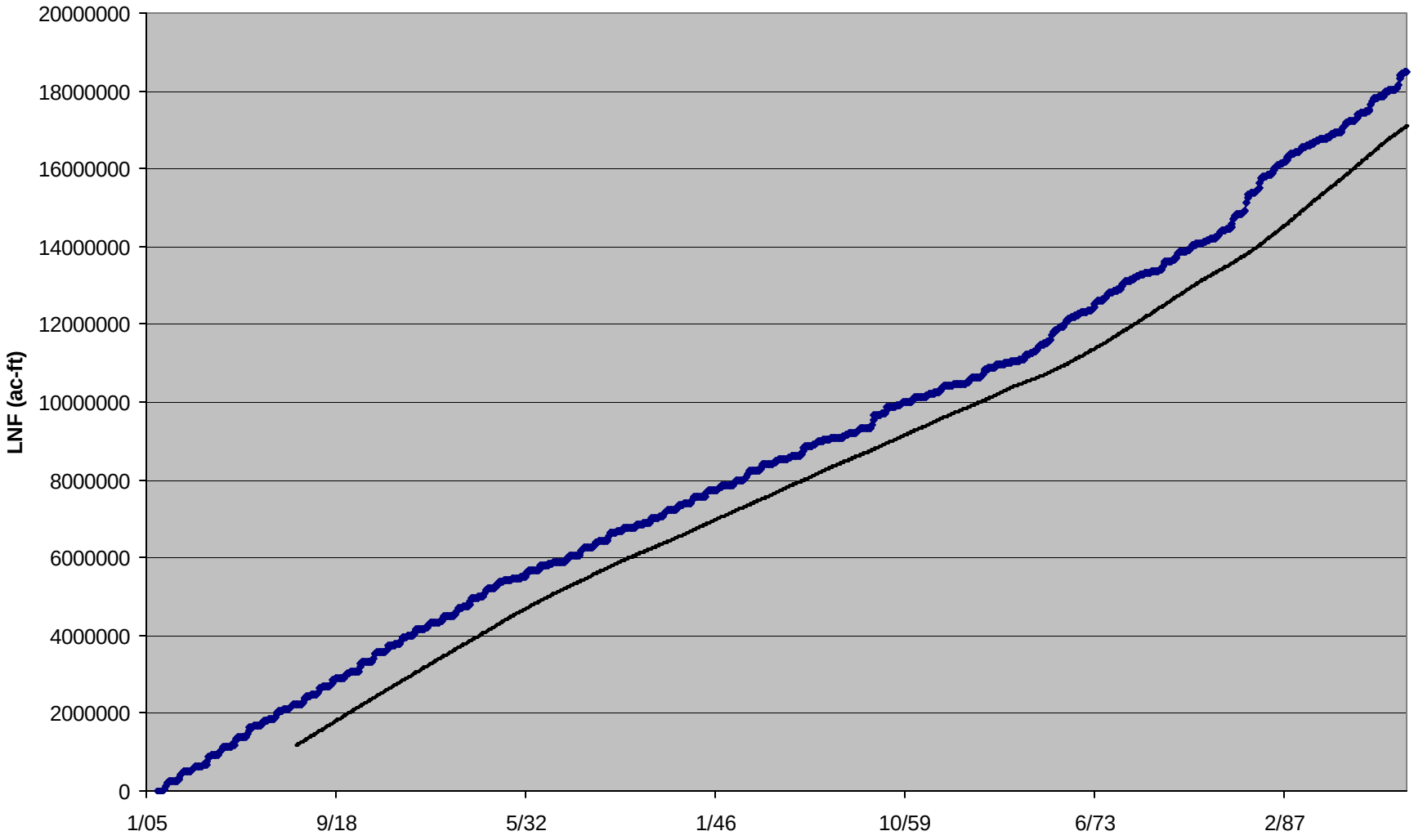
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 3



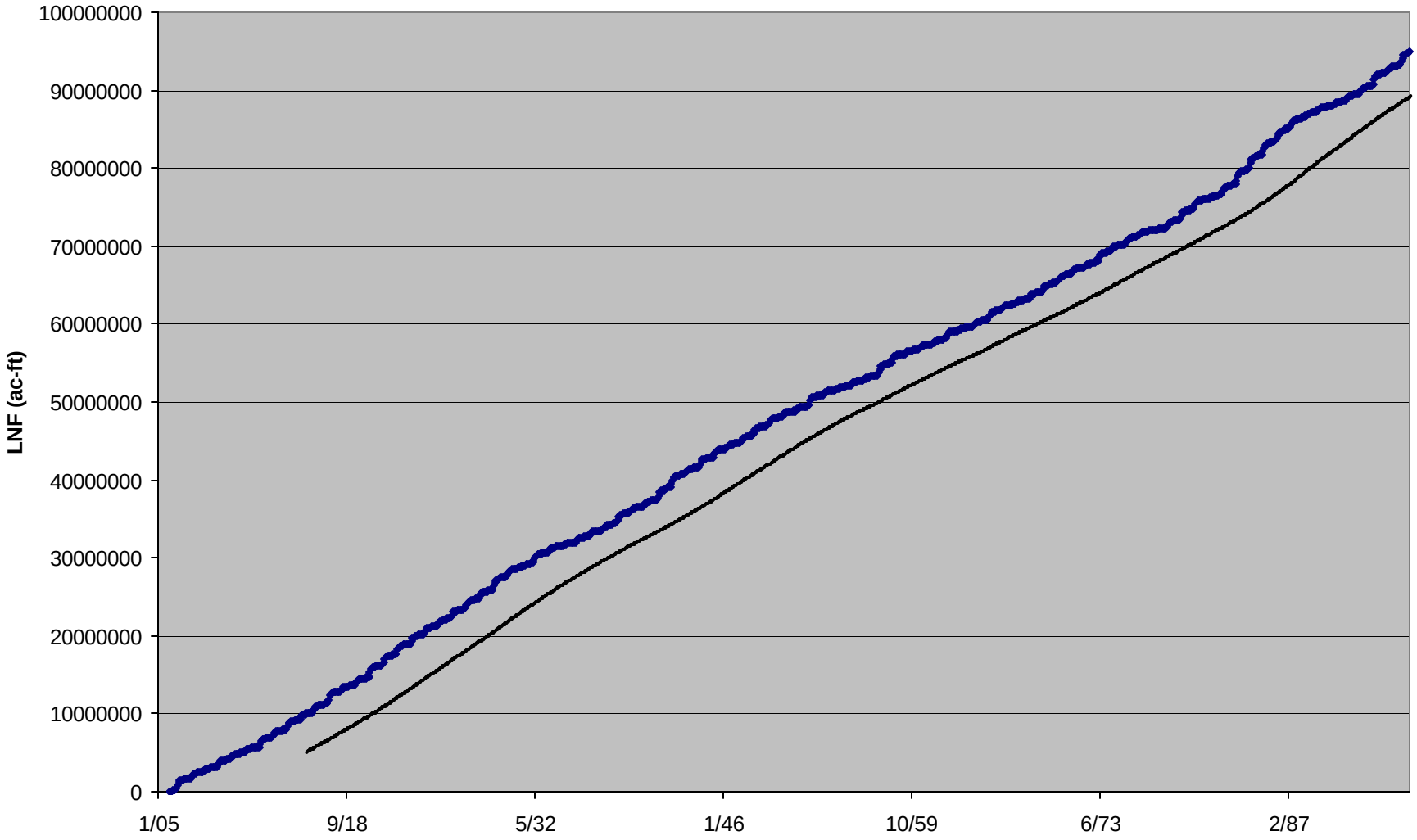
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 4



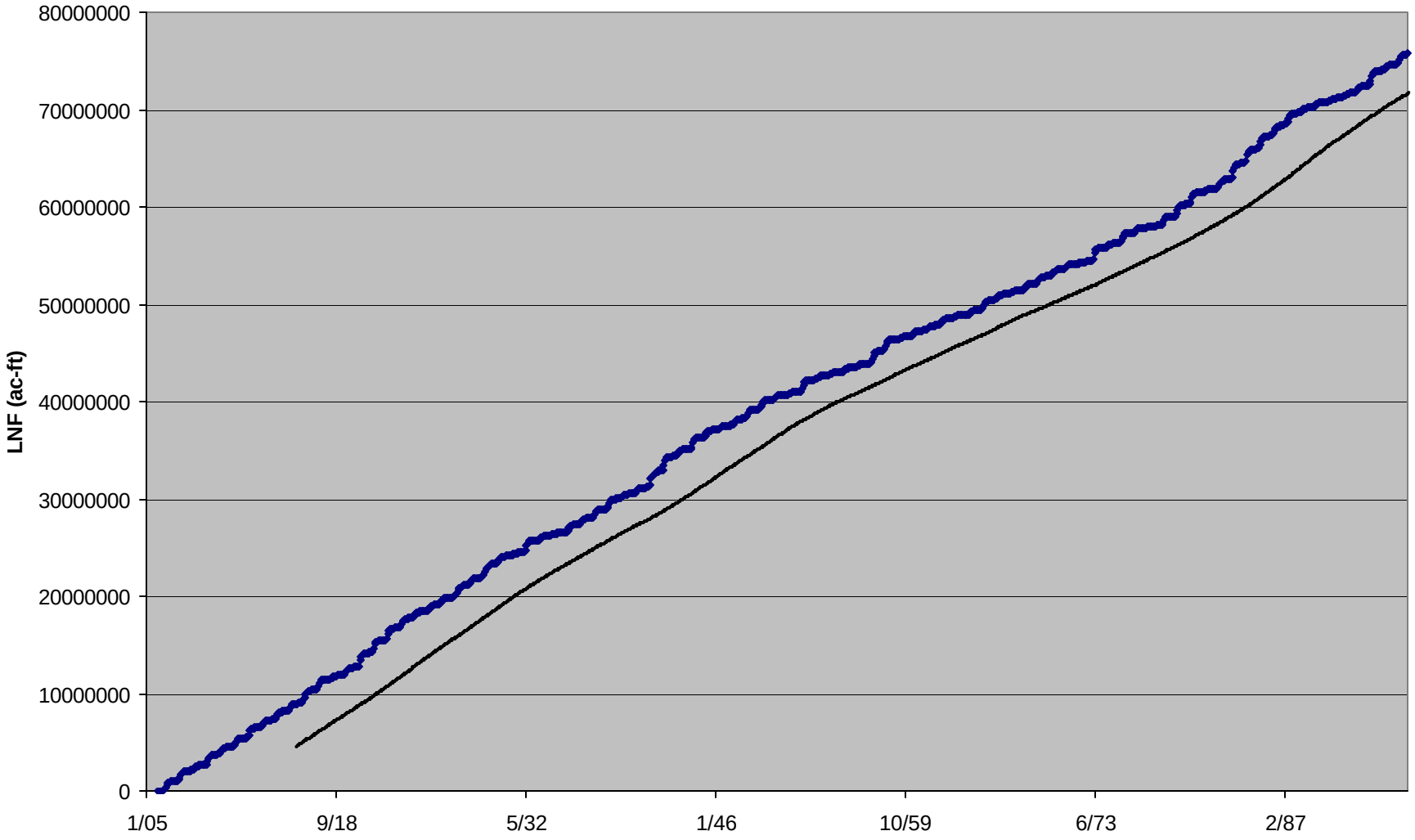
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 5



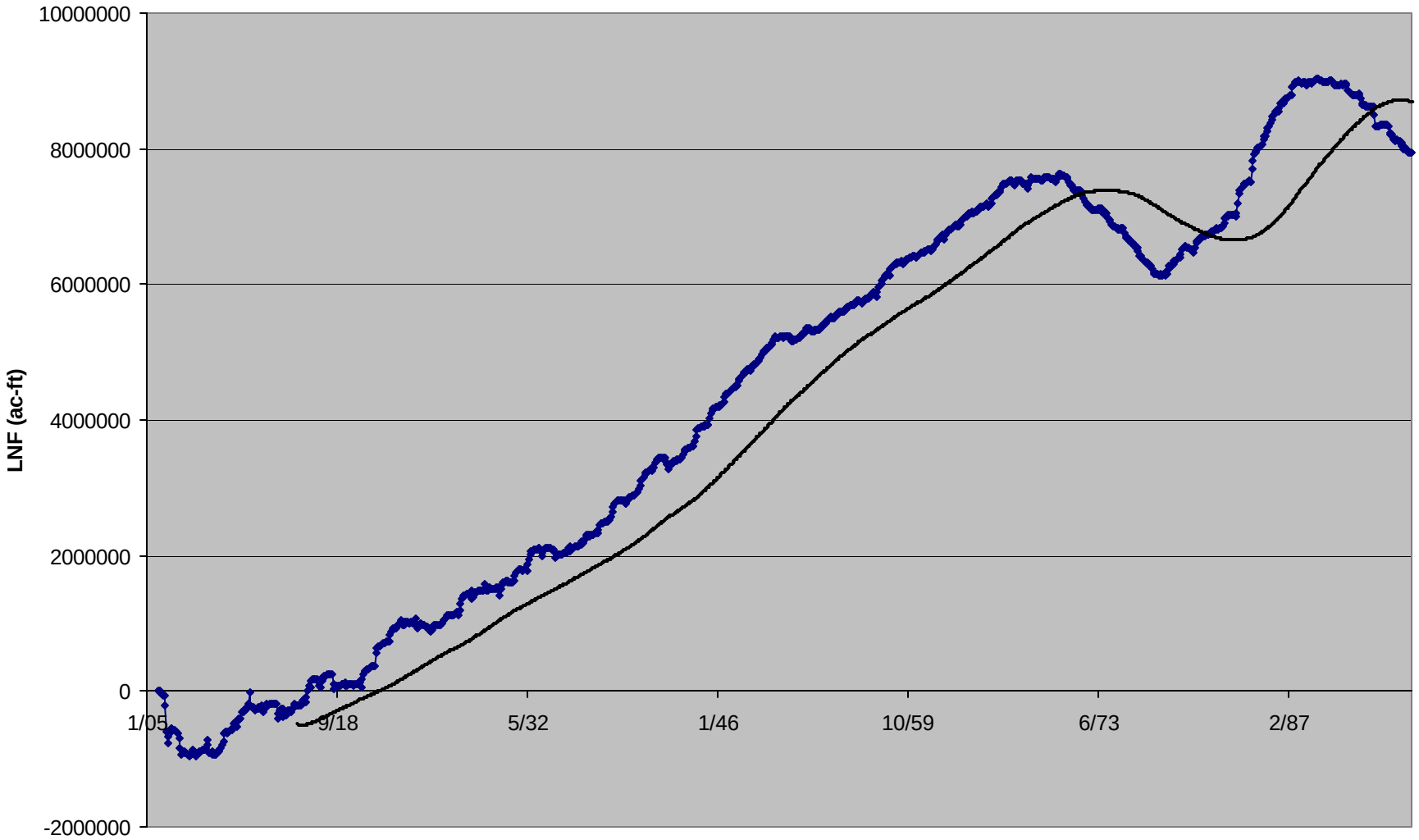
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 6



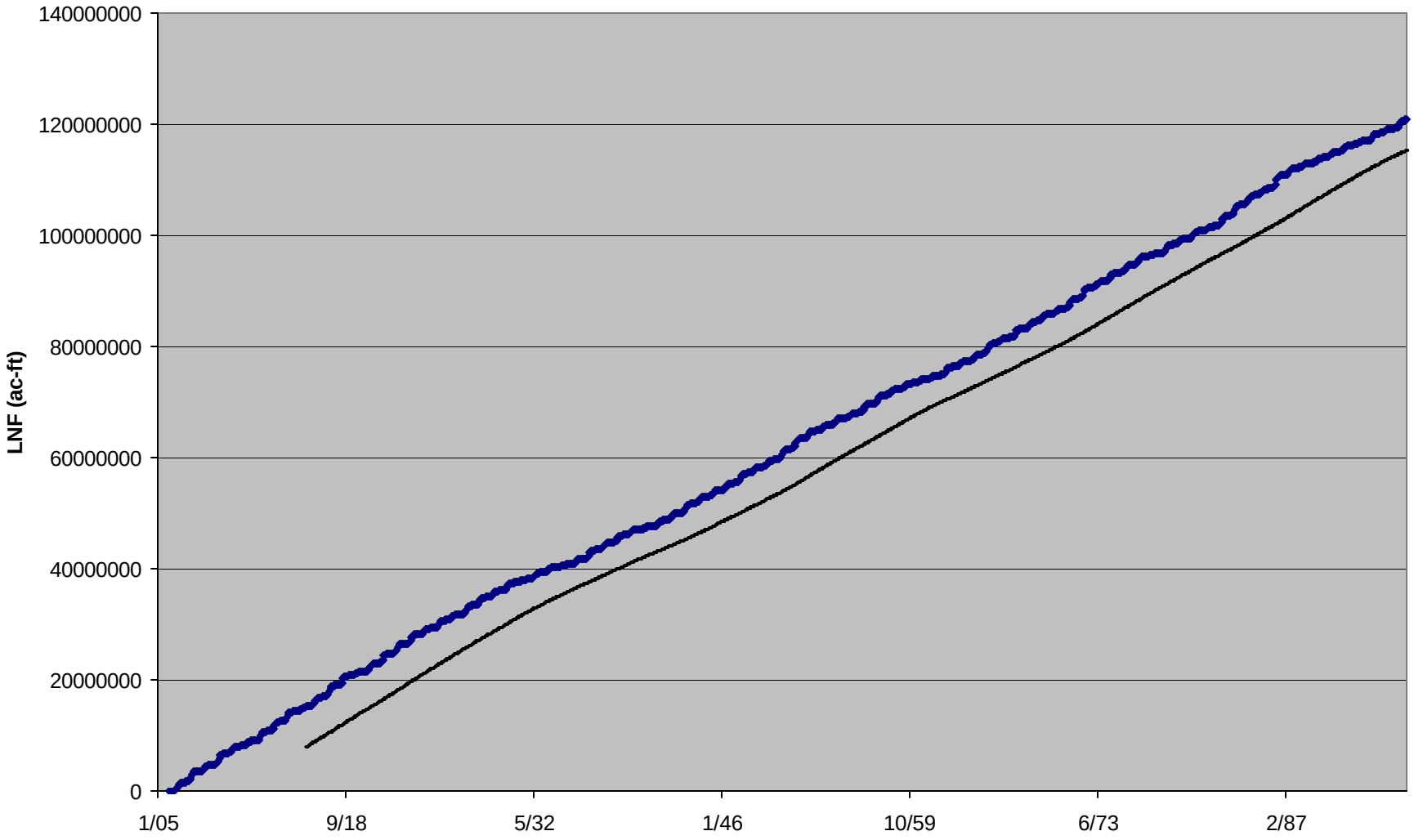
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 7



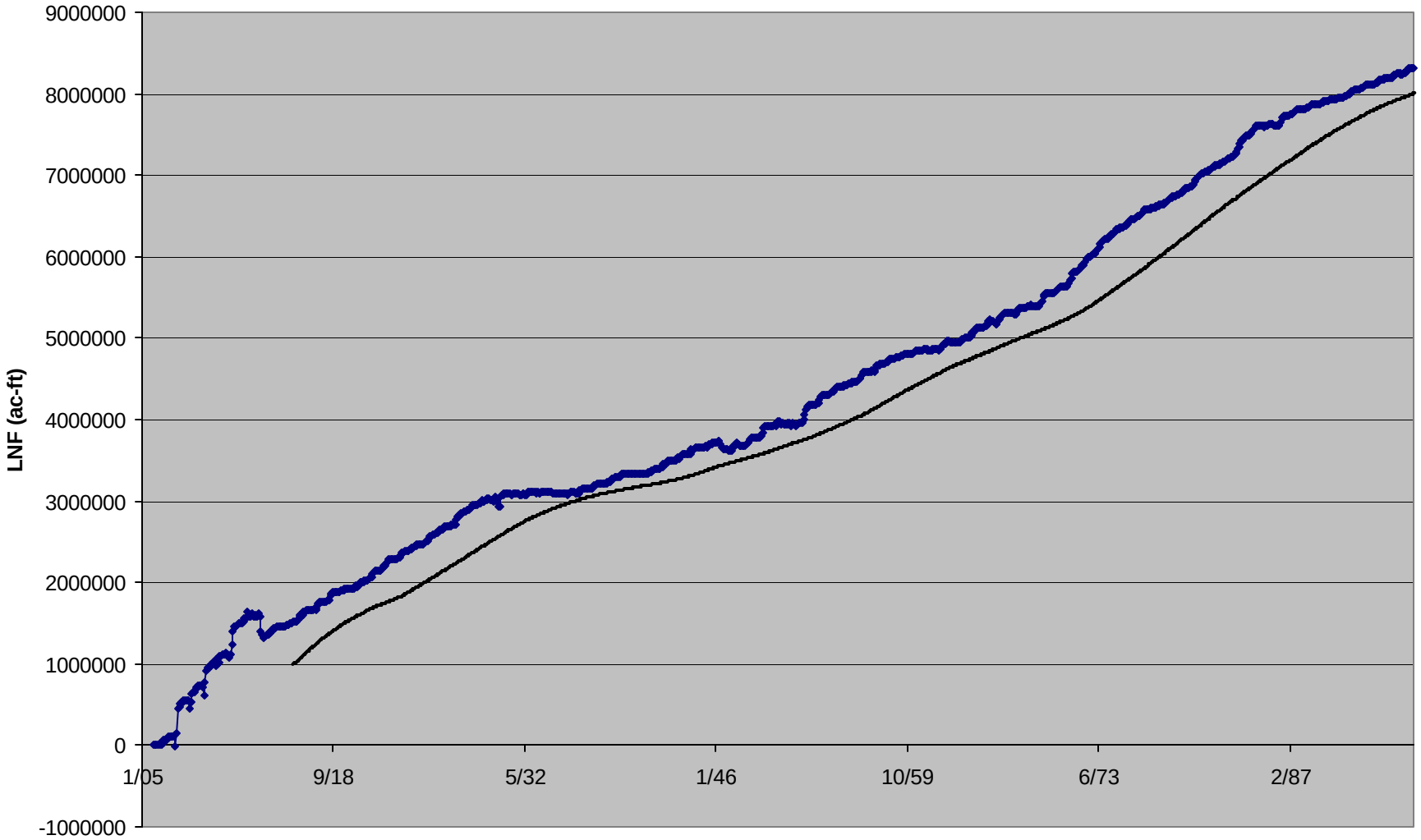
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 8



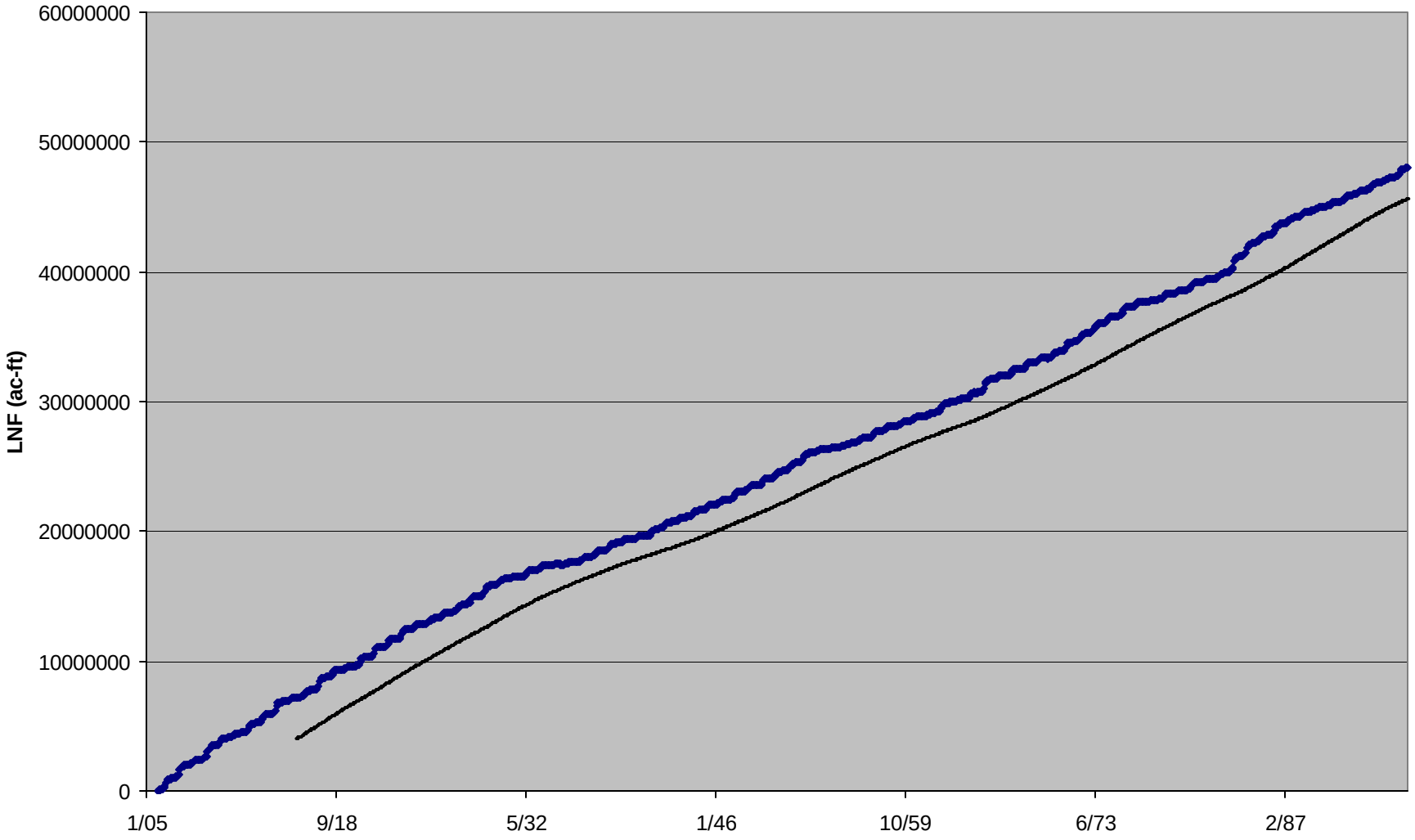
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 9



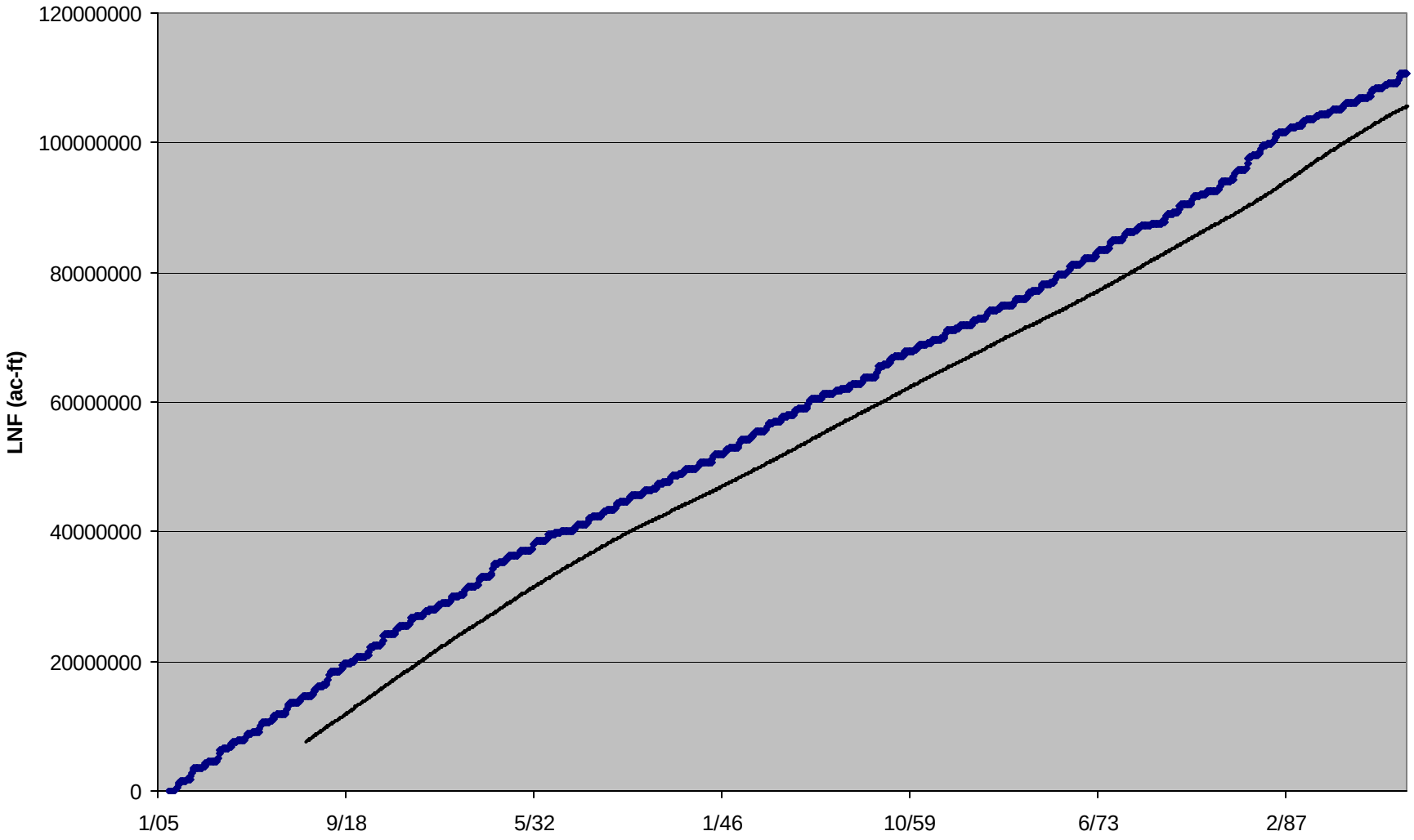
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 10



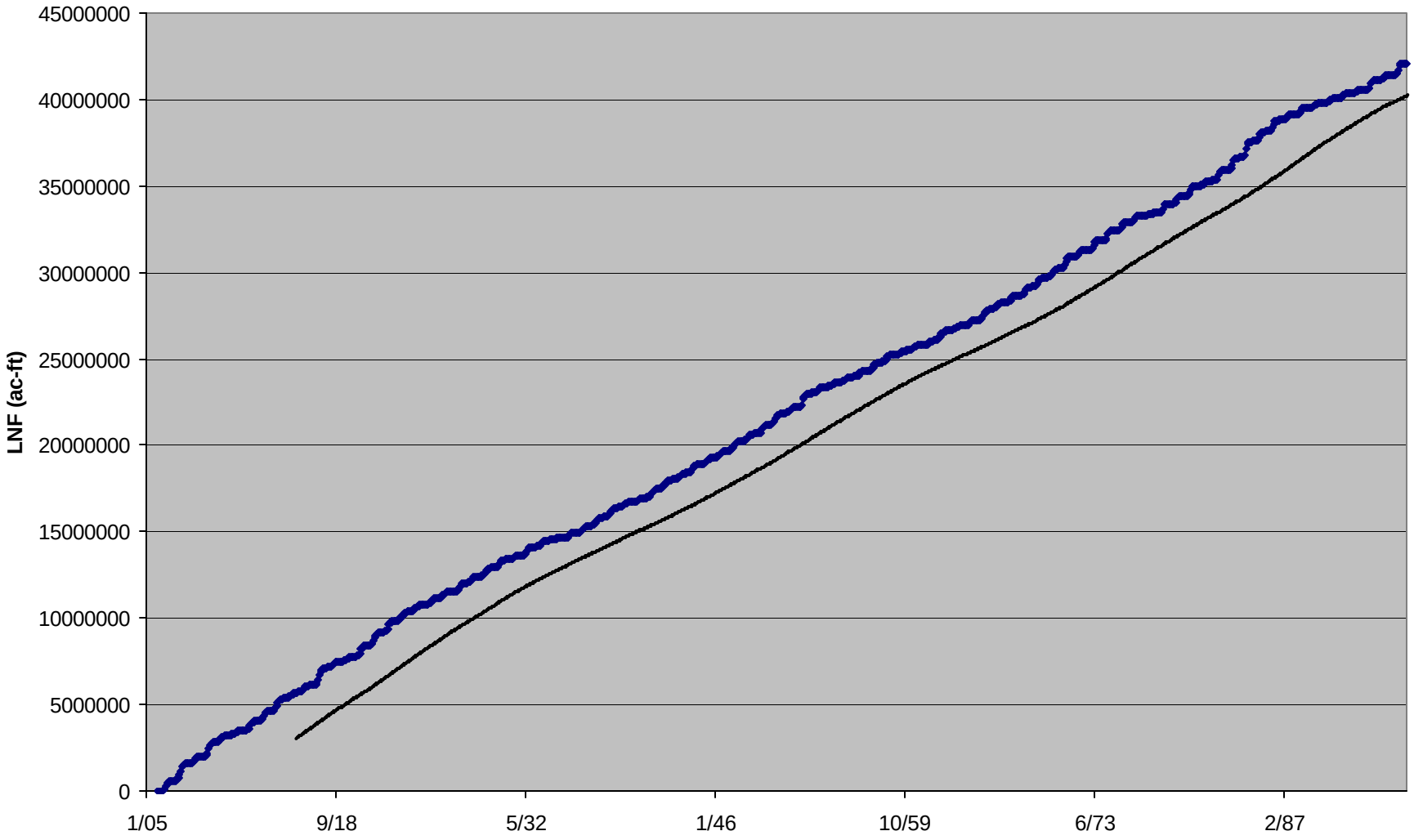
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 11



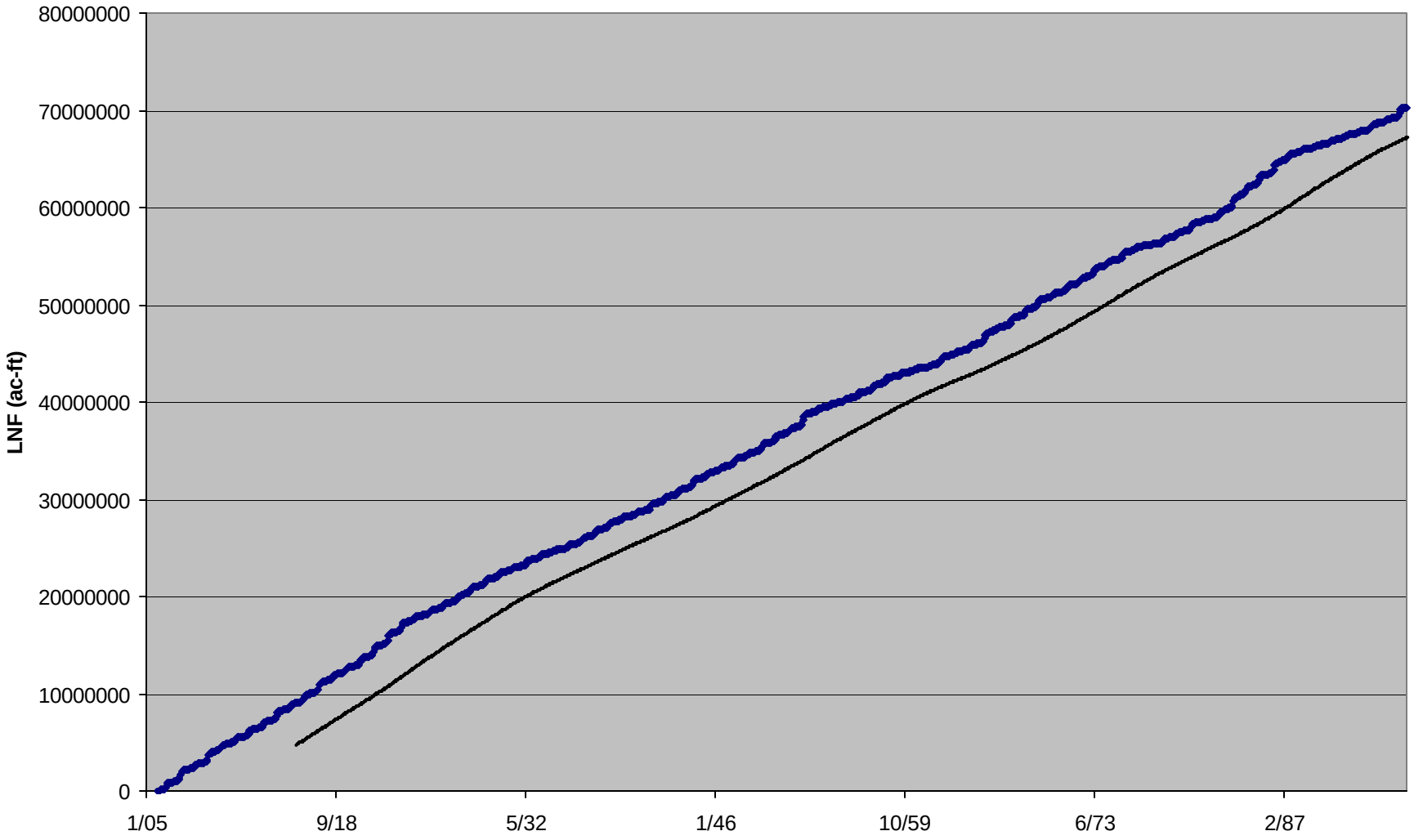
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 12



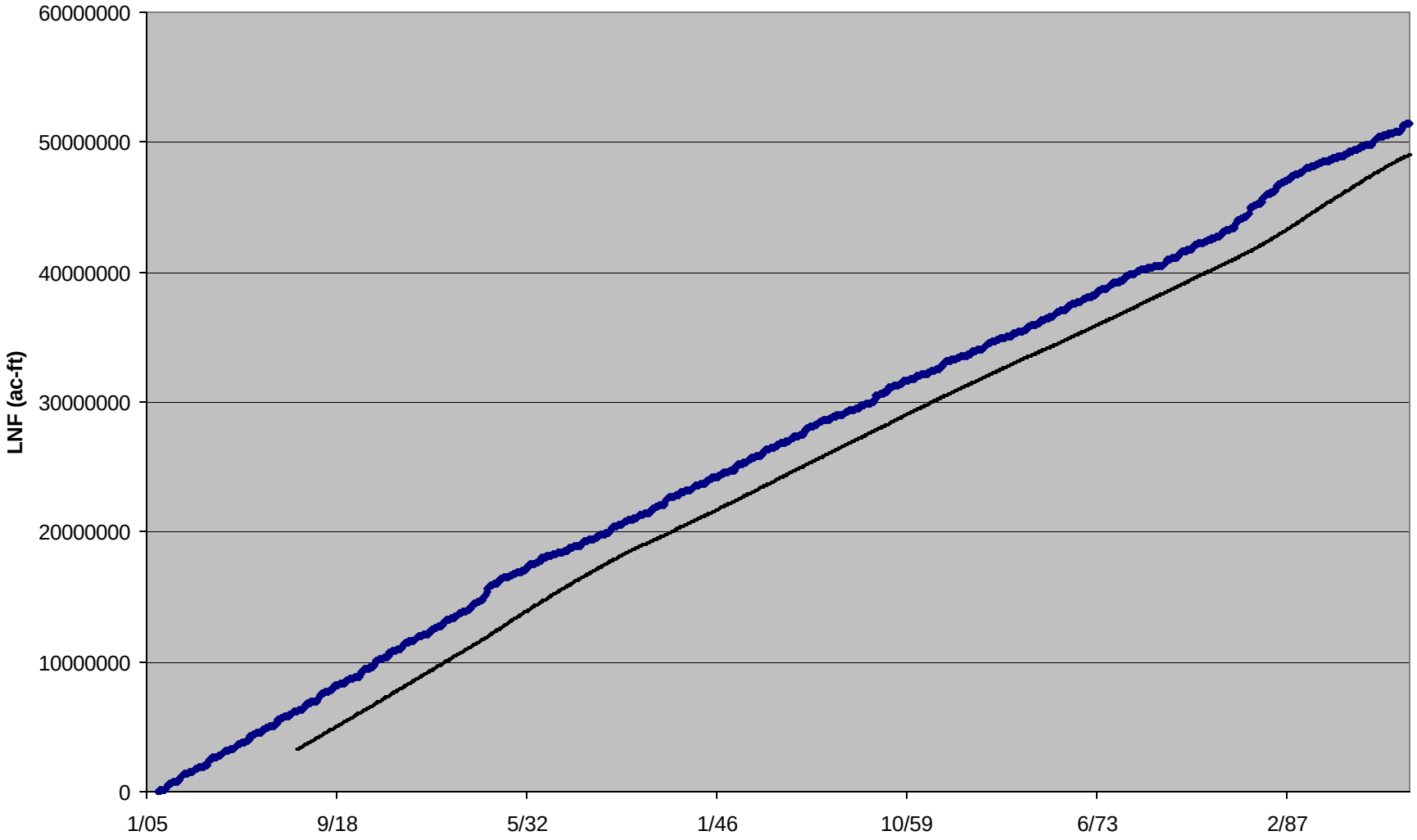
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 13



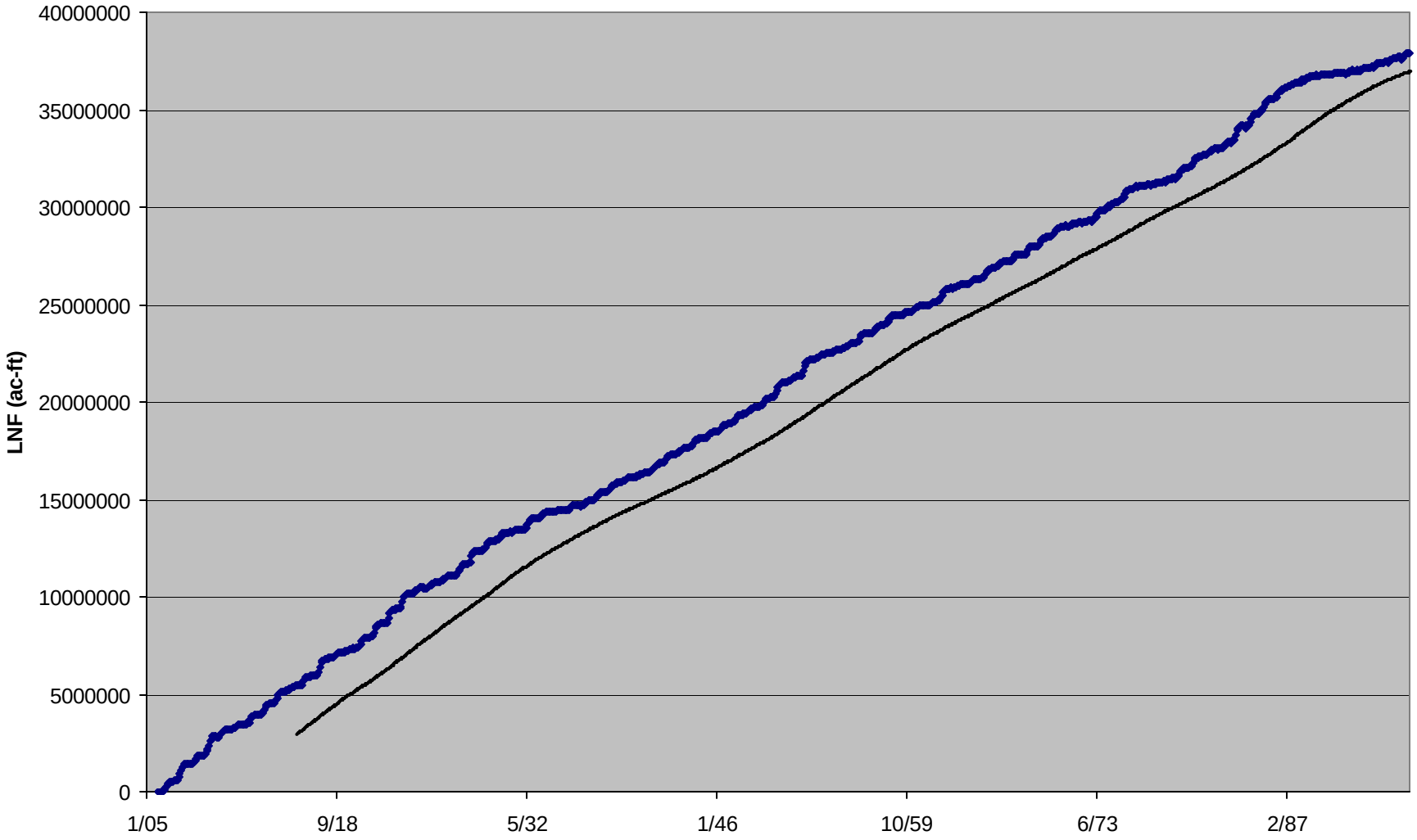
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 14



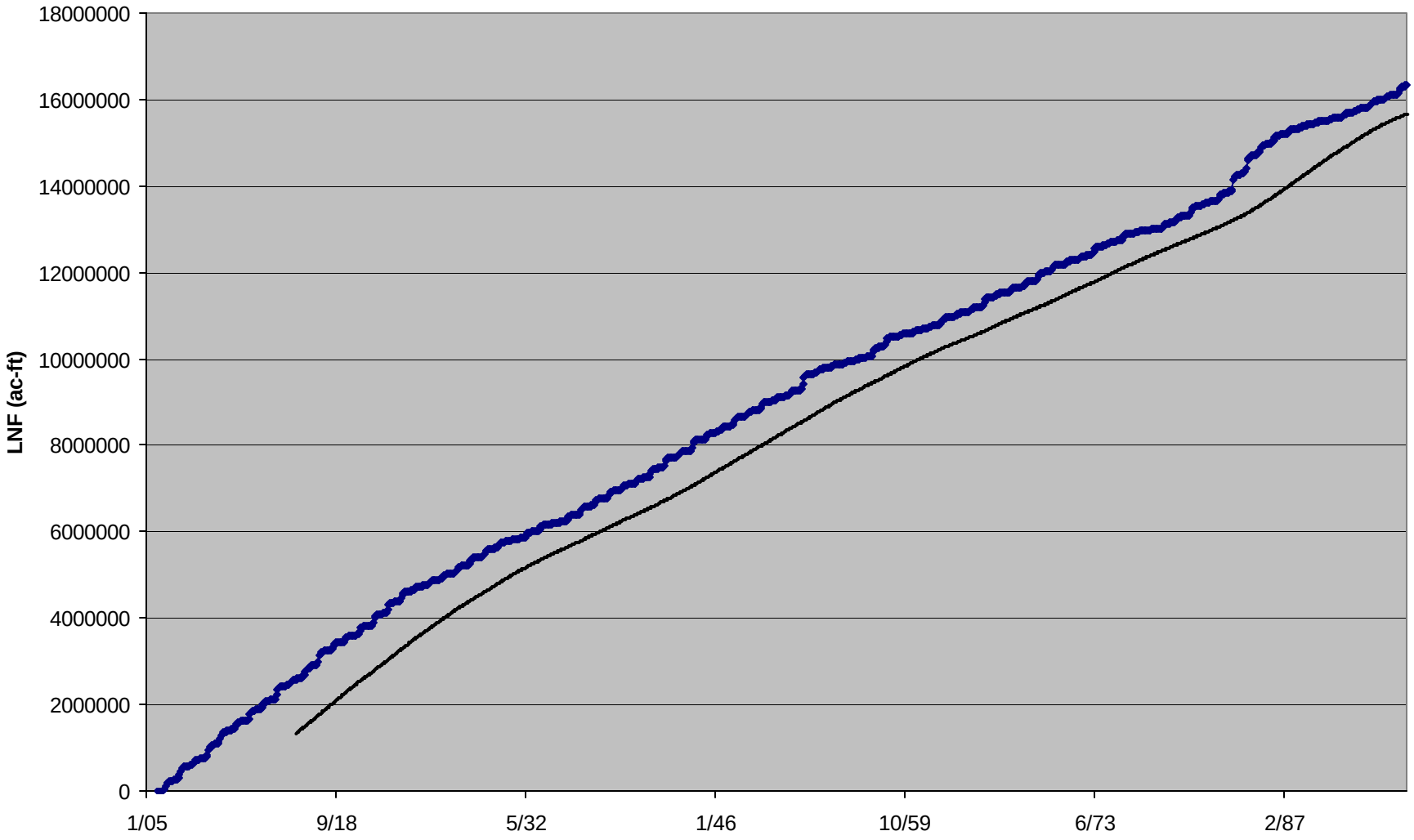
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 15



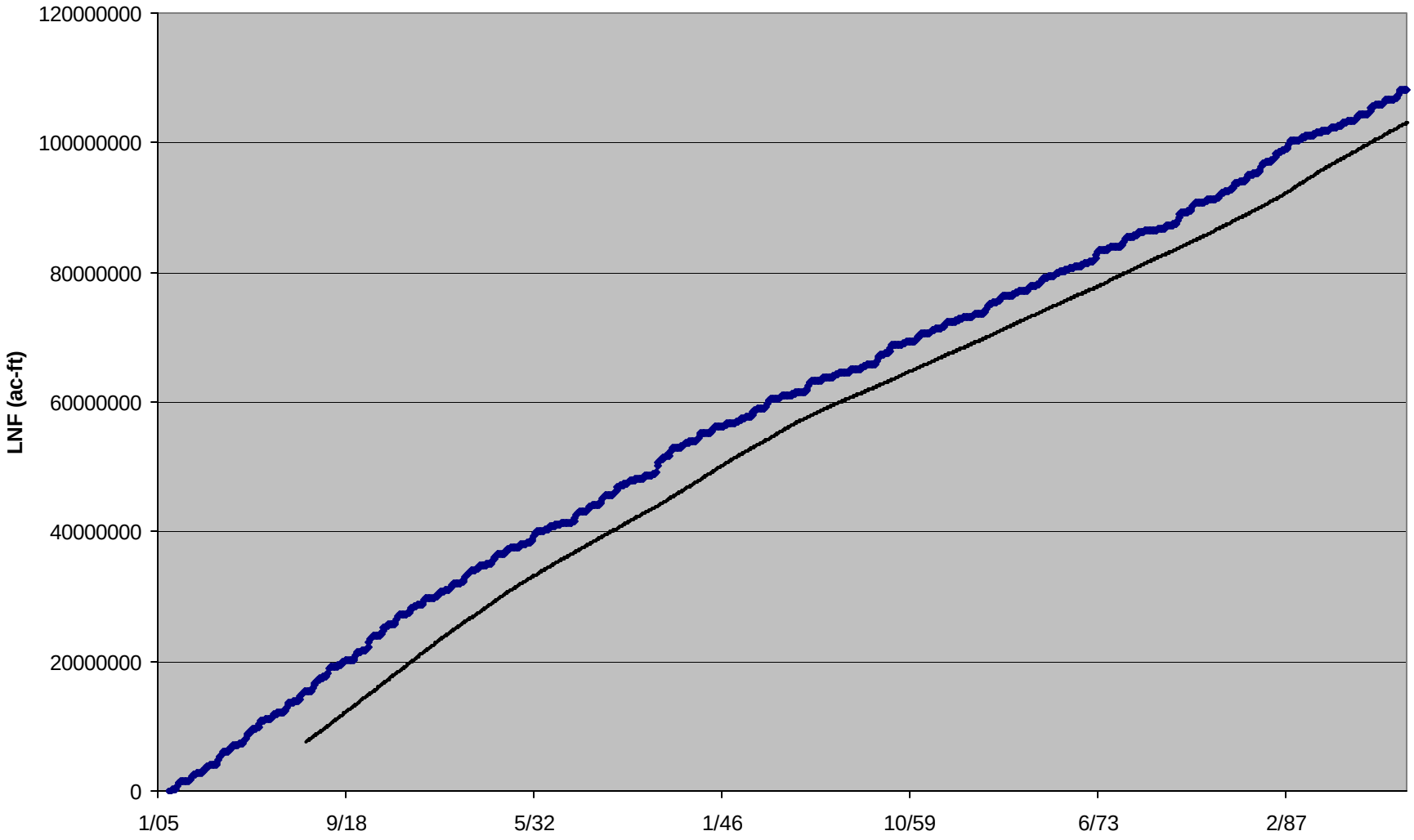
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 16



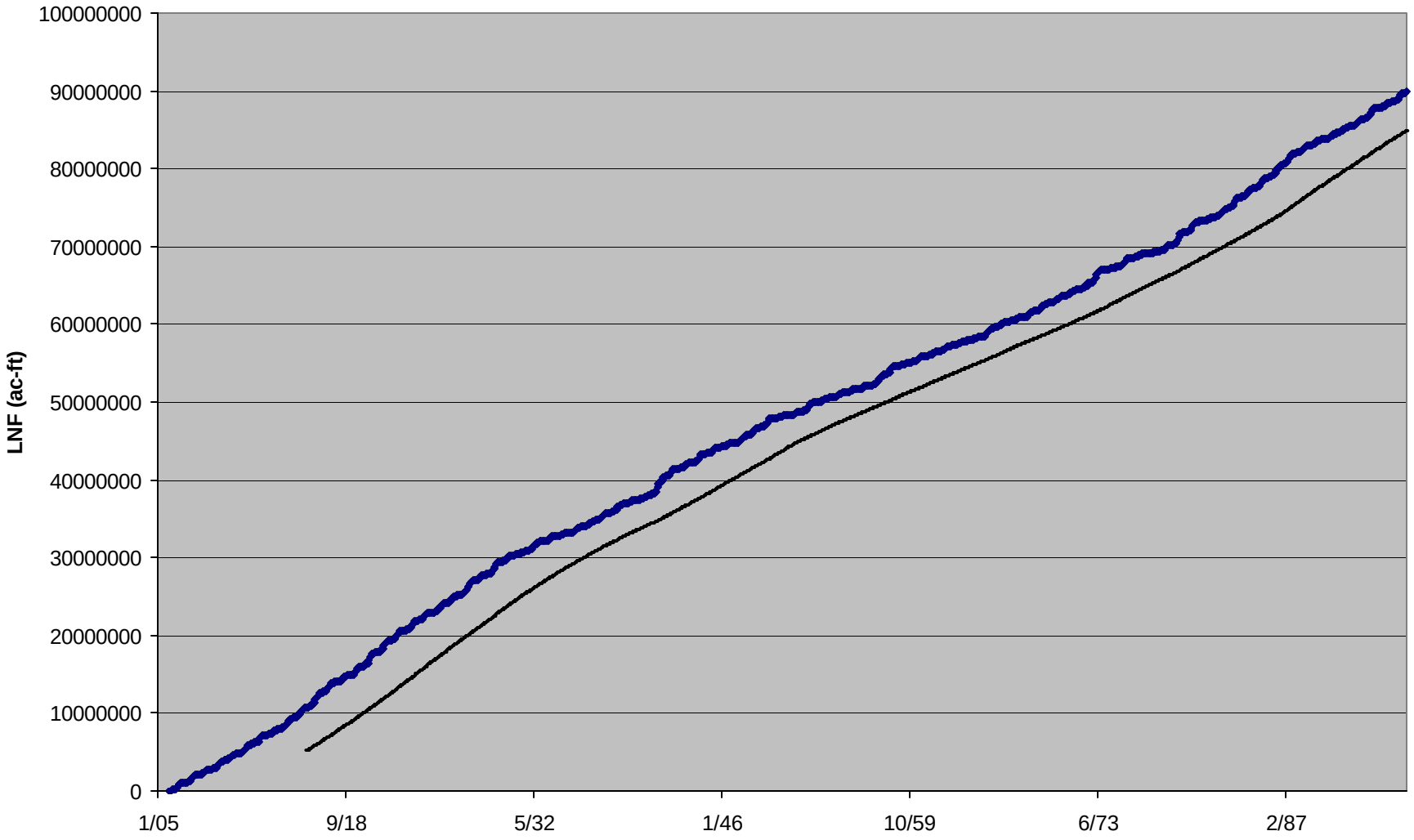
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 17



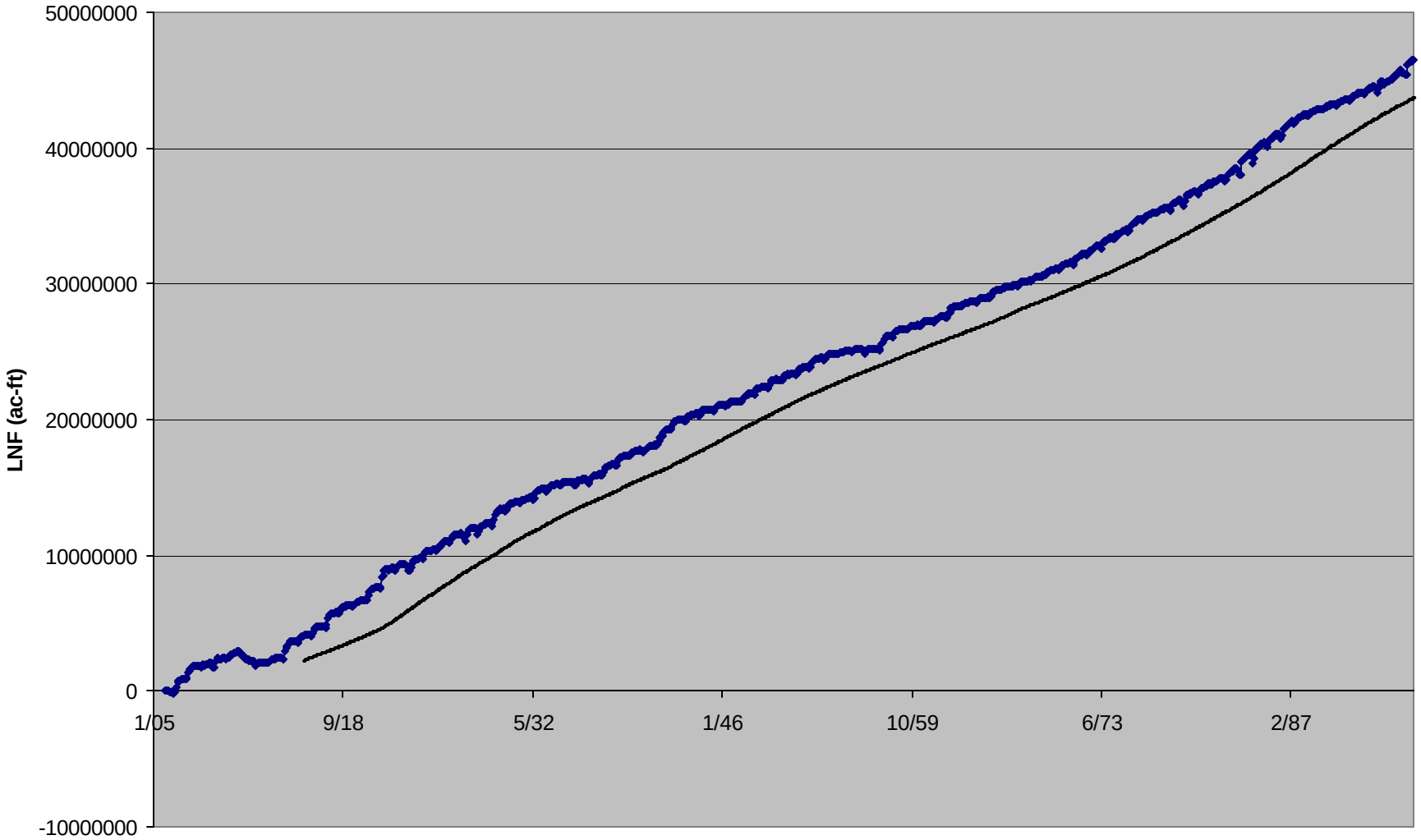
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 18



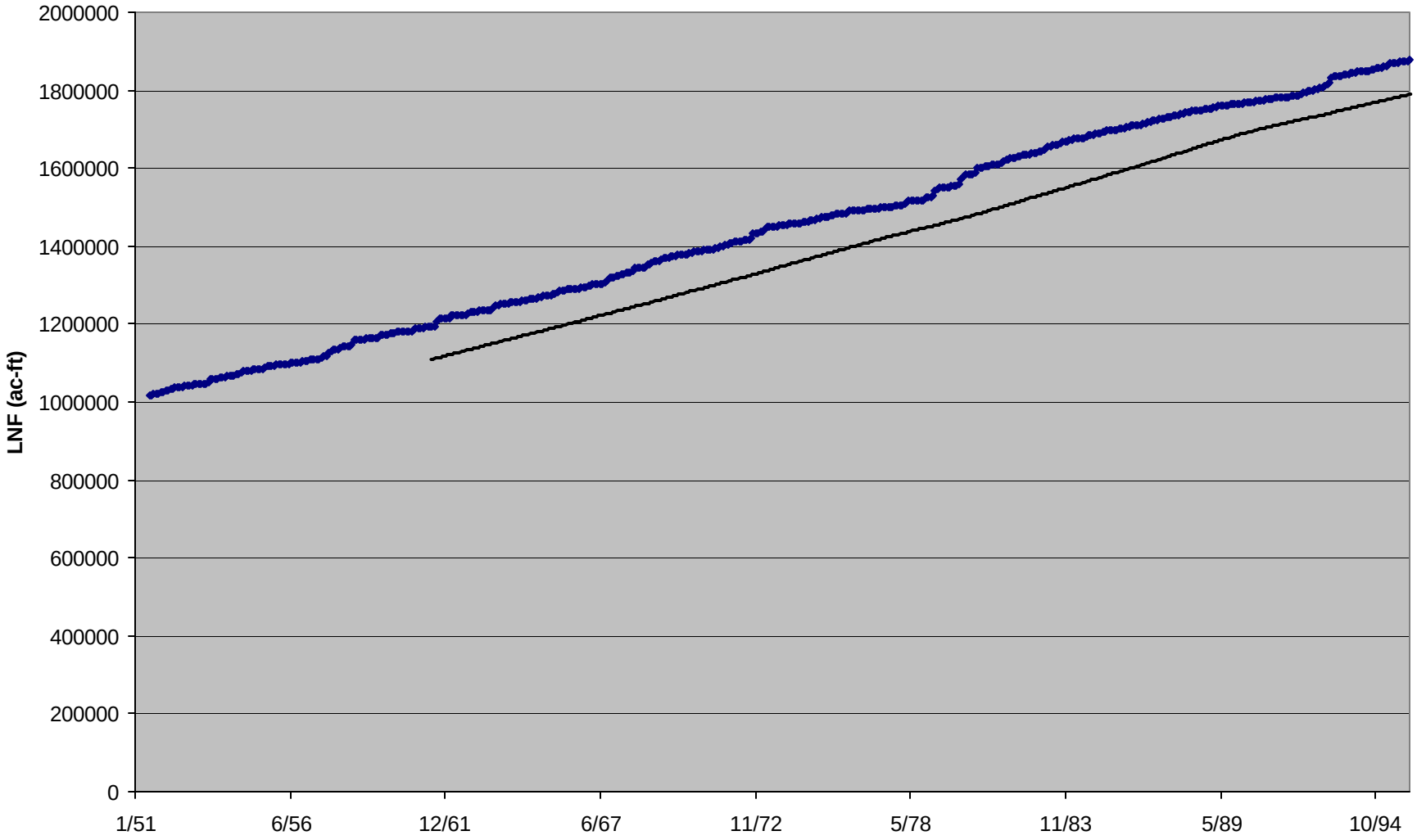
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 19



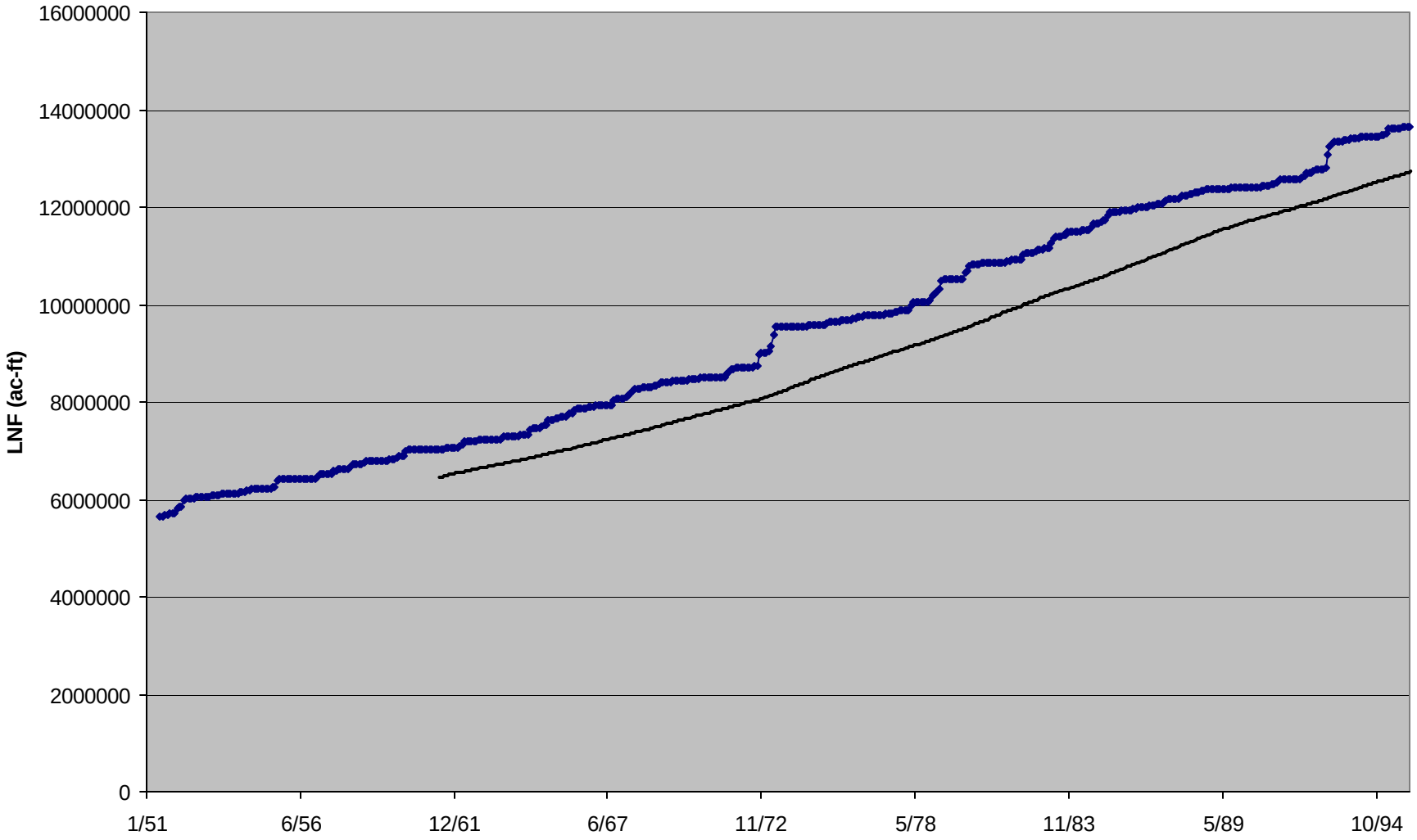
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 20



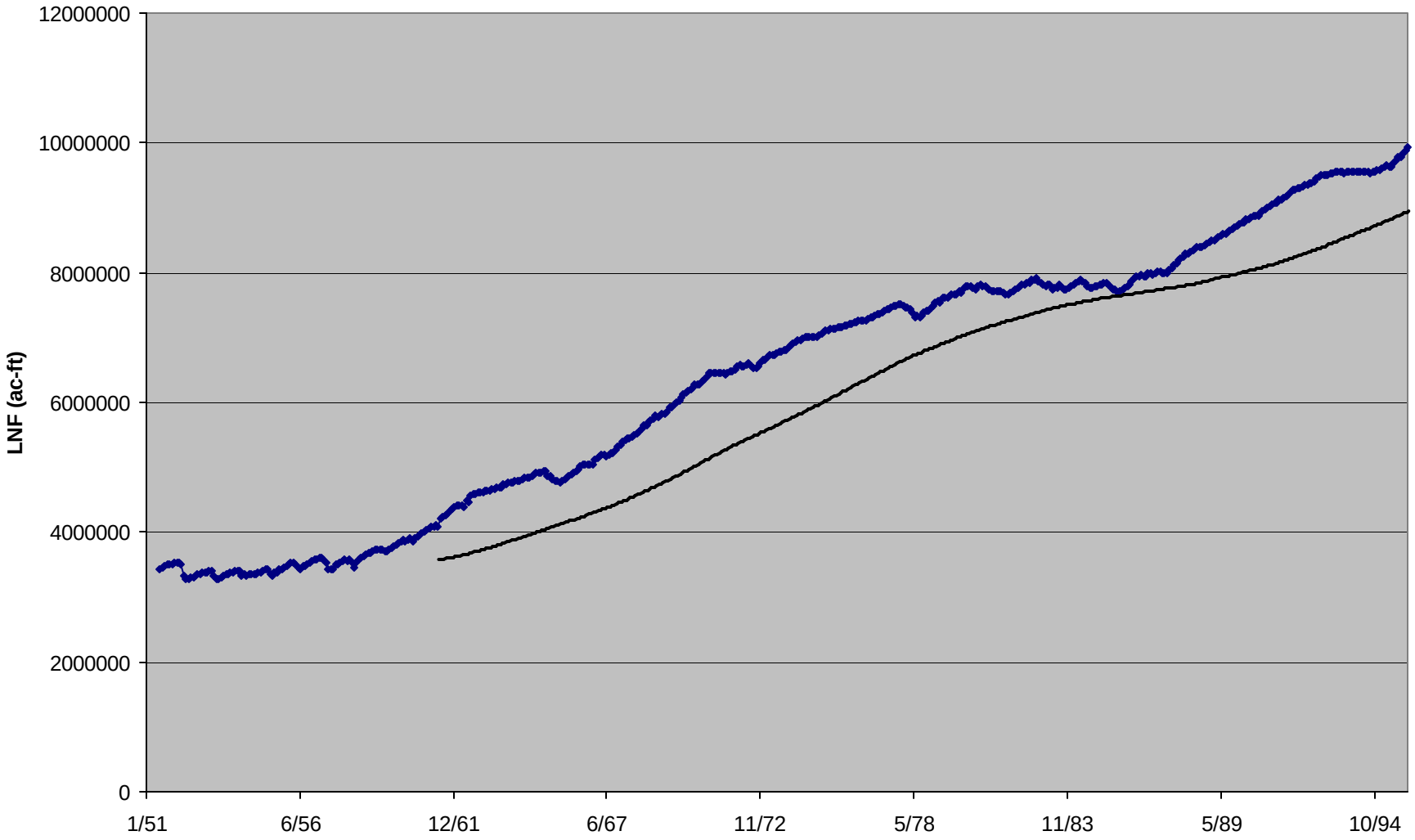
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 21



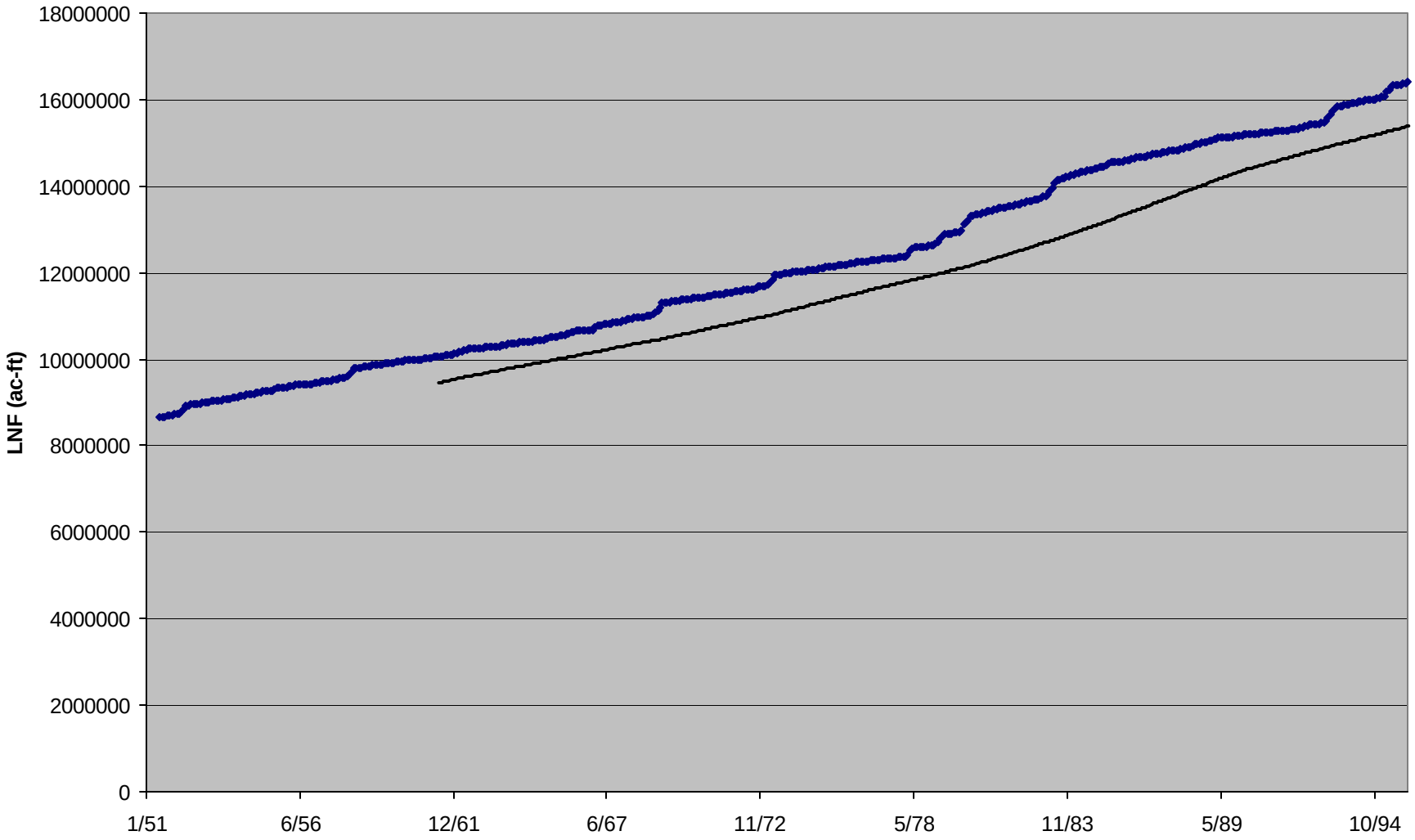
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 22



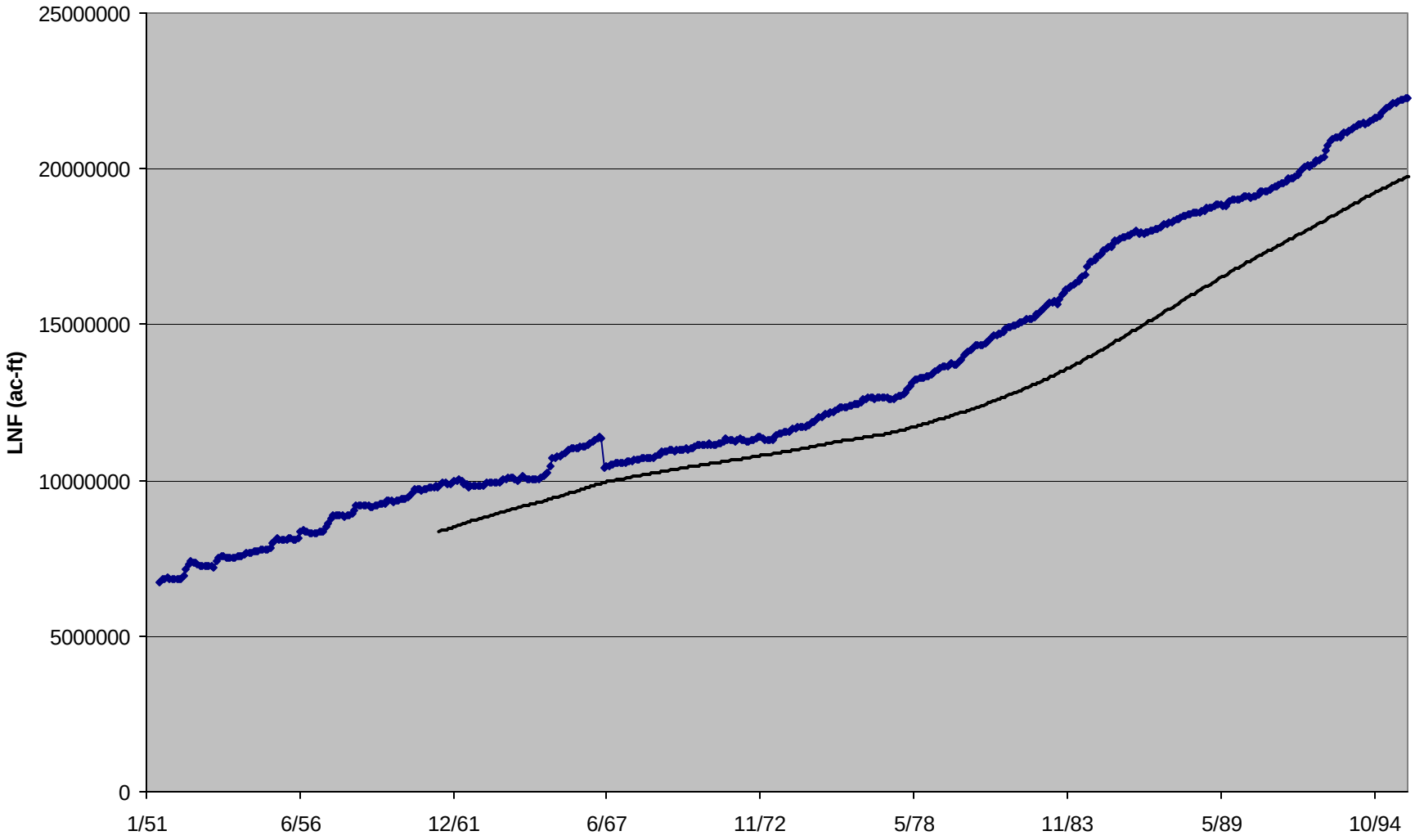
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 23



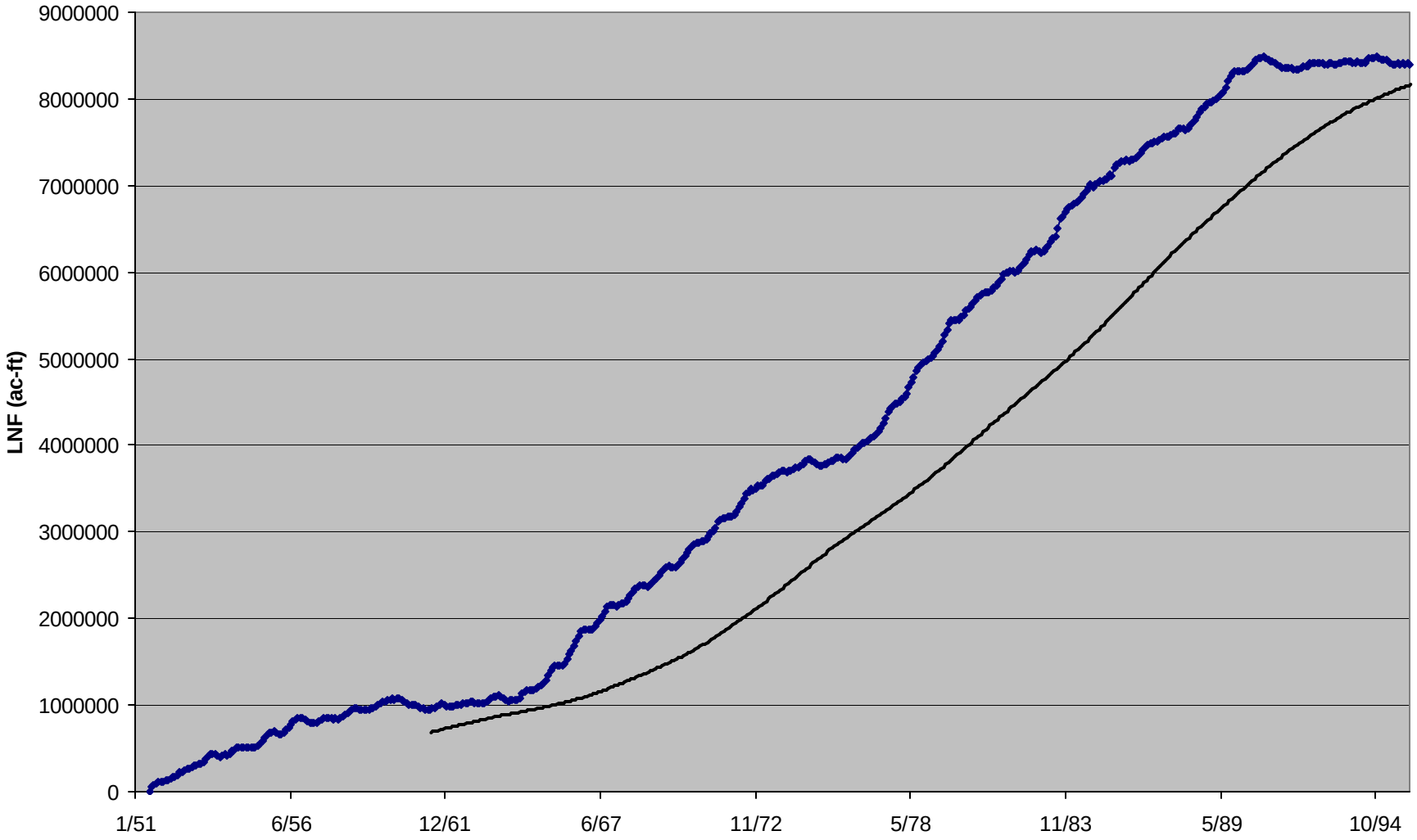
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 24



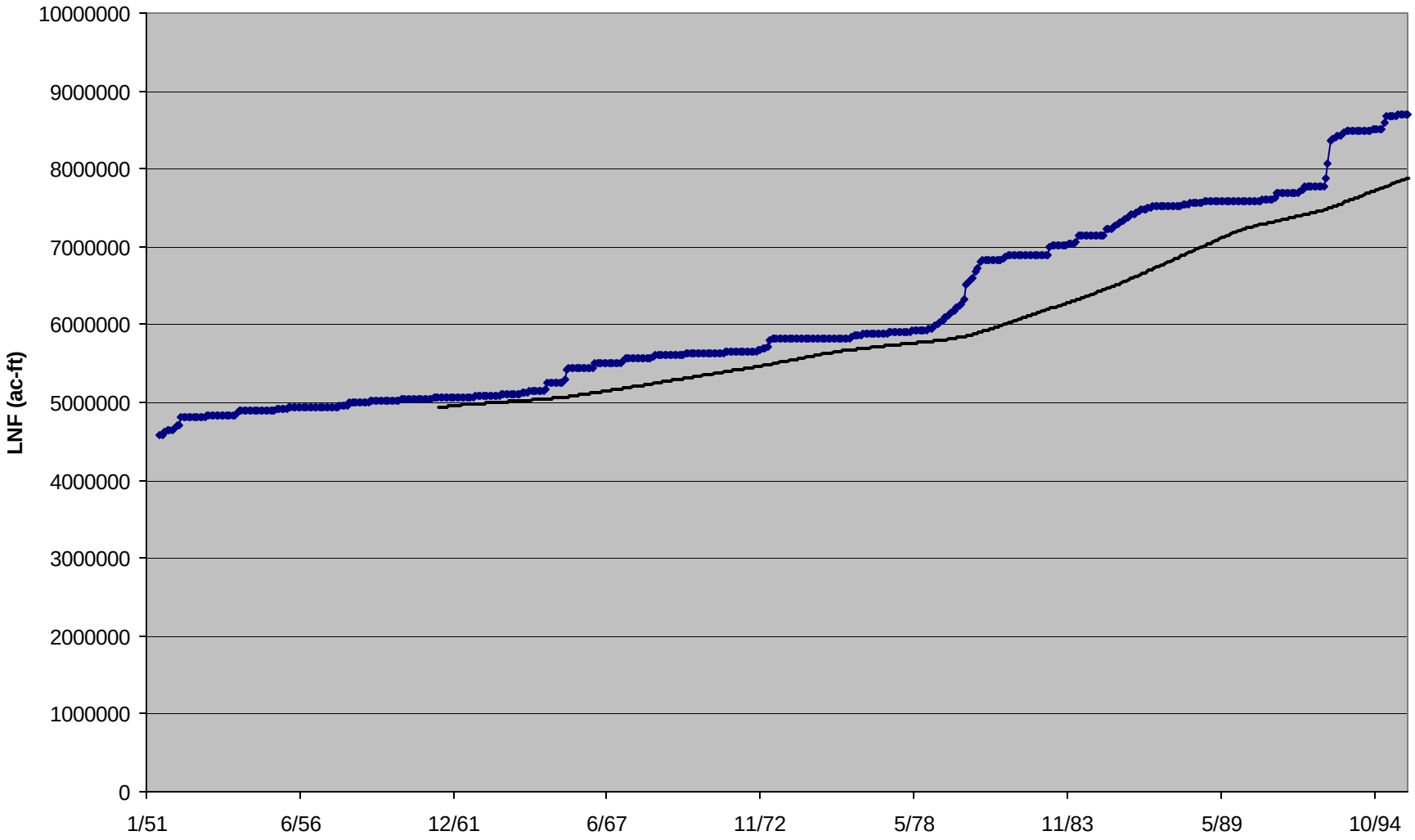
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 25



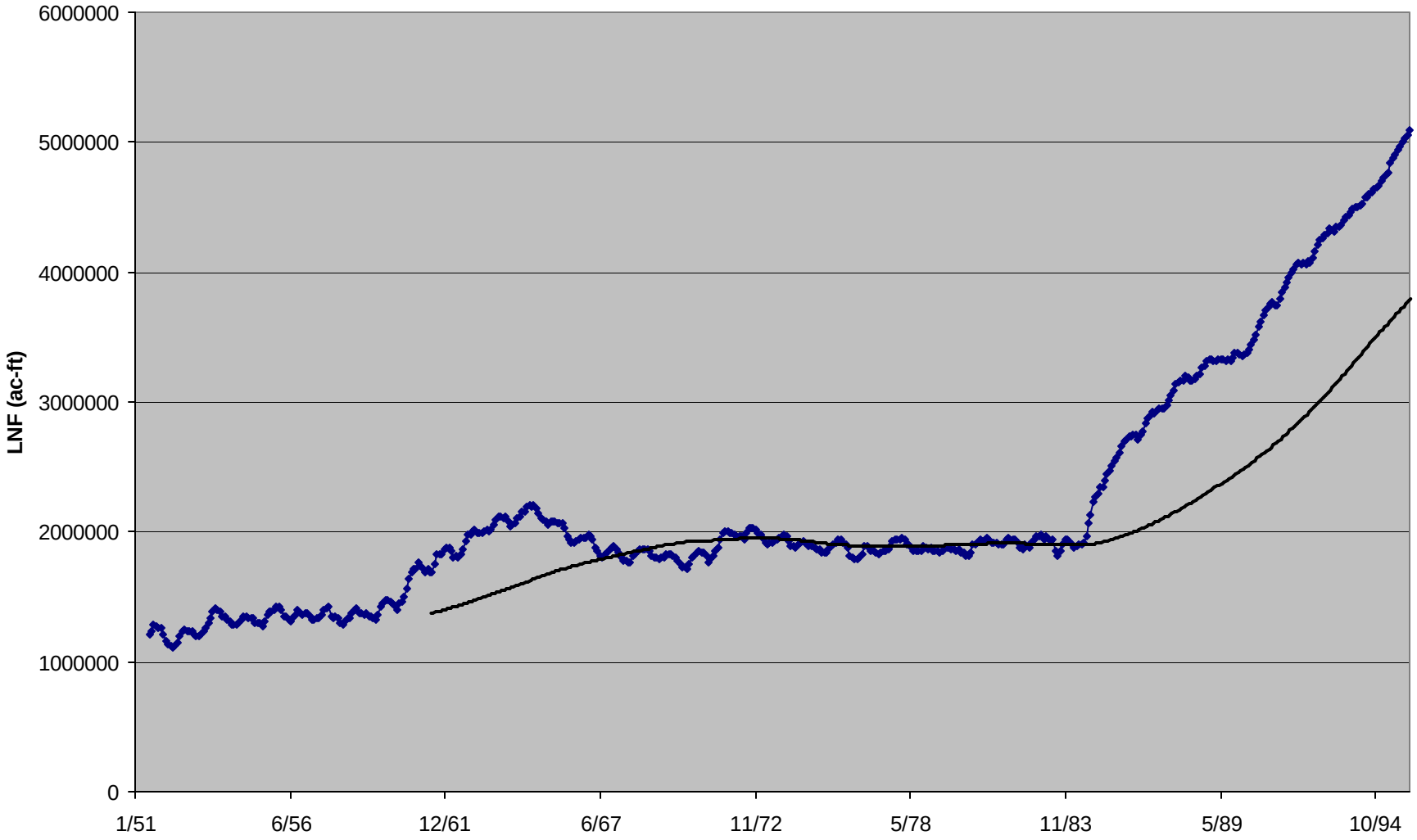
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 26



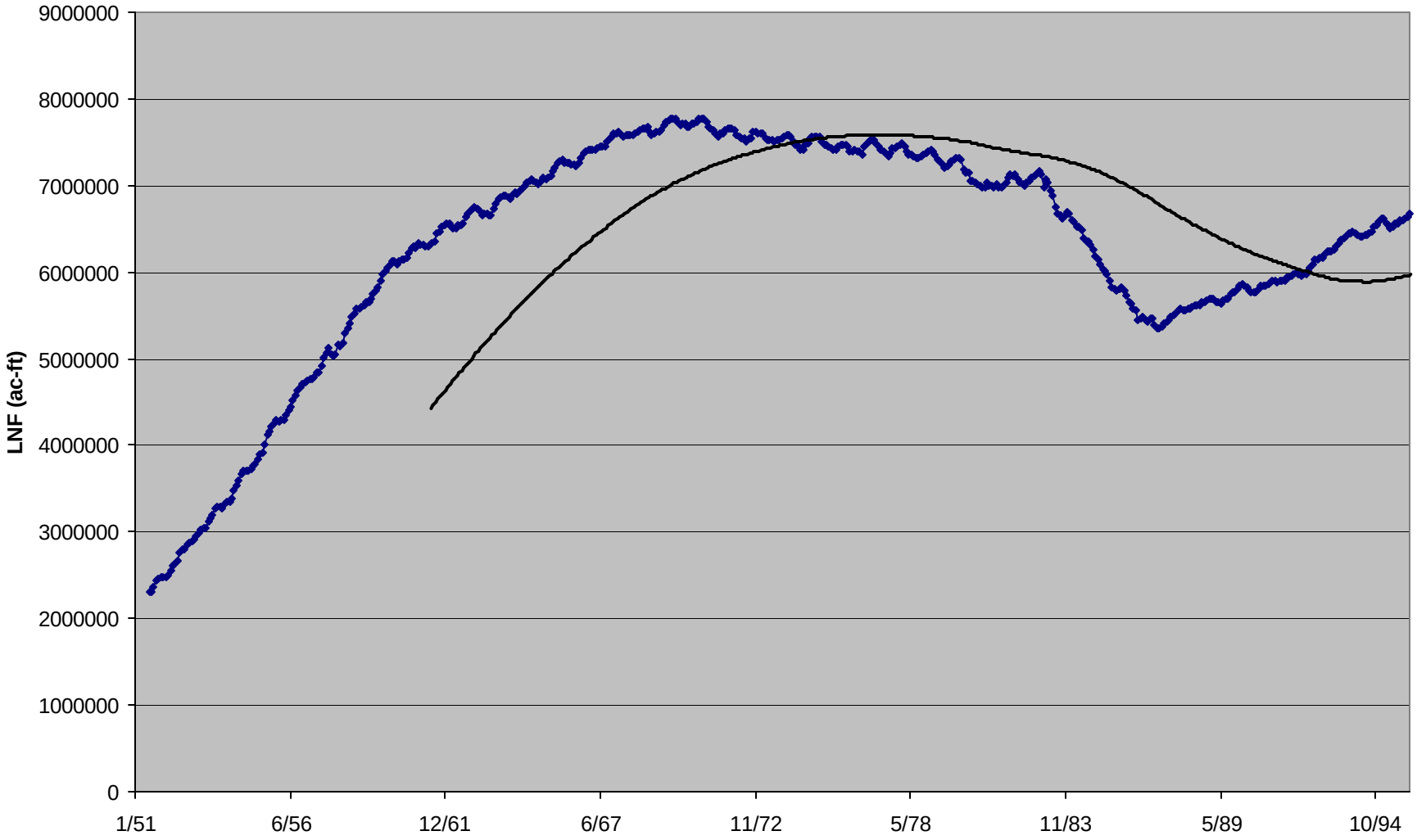
Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 27



Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 28



Simple Mass Curve (Local Flows, with 10-yr moving average)
Site 29



Example of non-relationships in Local Natural Flow

Local Inflow Non-Relationship Example