

unesp  **UNIVERSIDADE ESTADUAL PAULISTA**

Faculdade de Ciências e Tecnologia
Programa de Pós-Graduação em Ciências Cartográficas

CLÉLIA FRANCO



**VISÃO SISTÊMICA DO SÍTIO ARQUEOLÓGICO
PIRACANJUBA: A DESCOBERTA DE CONHECIMENTO
EM SÍTIOS ARQUEOLÓGICOS**

Presidente Prudente

2007

CLÉLIA FRANCO

**VISÃO SISTÊMICA DO SÍTIO ARQUEOLÓGICO
PIRACANJUBA: A DESCOBERTA DE CONHECIMENTO EM
SÍTIOS ARQUEOLÓGICOS**

Tese apresentado ao Programa de Pós-graduação em
Ciências Cartográficas da UNESP – Presidente Prudente
para obtenção do título de doutor.

Orientador: Prof. Dr. Nilton Nobuhiro Imai
Co-Orientadores: Prof. Dr. Neide Faccio Barrocá
Prof. Dr. Vilma Tachibana.

Presidente Prudente

2007

Franco, Clélia
F895 Visão sistêmica do Sítio Arqueológico Piracanjuba: a descoberta de conhecimento em sítios arqueológicos. / Clélia Franco. – Presidente Prudente: [s.n.], 2007
148 f. : il.

Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências e Tecnologia

Orientador: Nilton Nobuhiro Imai, Neide Faccio Barrocá, Vilma Tachibana

Banca: José Luiz de Moraes, Emília Mariko Kashimoto, Mário Hissamitsu Tarumoto, Milton Hirokazu Shimabukuro

Defesa: 26/02/2007

1. Descoberta de conhecimento. 2. Mineração de dados. 3. Análise de correspondência. 4. Estatística multivariada. 5. Modelo numérico de terreno. 6. Arqueologia. I. Imai, Nilton Nobuhiro II. Barroca, Neide Faccio. III. Tachibana, Vilma. IV. Banca. V. Título.

CDD (18.ed.) 623.71

AGRADECIMENTOS

É difícil escrever a página de agradecimentos de uma tese, expressar sentimentos fortes com palavras que não conseguem traduzir a gratidão por ter sido merecedor da confiança e do apoio das muitas pessoas envolvidas neste processo.

A todos que colaboraram direta ou indiretamente na elaboração deste trabalho e, em especial:

à minha família, que soube respeitar e entender as longas horas de estudo solitário;

aos meus orientadores Profs. Drs. Nilton Nobuhiro Imai, Neide F. Barrocá e Vilma M. Tachibana que souberam estimular, aconselhar, criticar e orientar de forma segura este trabalho sem cercear a minha liberdade e maneira de trabalhar;

ao Prof. Dr. José Luiz de Moraes, diretor do MAE/USP e arqueólogo responsável pelo trabalho de pesquisa na área do Sítio Arqueológico Piracanjuba, por disponibilizar os dados de pesquisa de sua equipe no referido Sítio, demonstrando desta forma o seu entendimento de que a Ciência não é feita de segmentos estanques, mas sim de segmentos que se unem em um objetivo comum, a ampliação do conhecimento;

aos meus colegas, companheiros de estudos e também responsáveis pelos momentos de descontração e alegria que dão ânimo e revigoram as forças para prosseguir no caminho.

Que Deus abençoe a todos e **MUITO OBRIGADO!**

RESUMO

Nas últimas décadas, a capacidade de gerar e coletar dados aumentou rapidamente, gerando a necessidade do desenvolvimento de novas técnicas e ferramentas capazes de processar e analisar esses dados descobrindo informações novas e úteis. Surgindo um proeminente campo de pesquisa para a extração de conhecimento de dados – Descoberta de Conhecimento em Banco de Dados.

Pela aplicação da metodologia da descoberta de conhecimento indireto aos atributos dos fragmentos cerâmicos coletados ao nível do solo no Sítio Arqueológico Piracanjuba – Piraju SP, este trabalho pretende prover aos peritos em arqueologia uma visão sistêmica capaz de auxiliá-los no conhecimento das populações pretéritas que ali habitaram.

Palavras chave: Descoberta de conhecimento; mineração de dados; estatística multivariada; análise de correspondência; modelo numérico de terreno; arqueologia.

ABSTRACT

In the last decades, the capacities to produce and collect data has grown fast and the development of news techniques and tools capable to processes and analyze this datas discovering new and useful information as necessary. Therefore, a huge research area has beginning for the extraction of data understanding – Knowledge Discovery in Database.

The indirect knowledge discovery applied to ceramic fragment collected at soil level in Piracanjuba's – Piraju, SP aims give to archaeology experts a whole vision able to be useful knowledge of the past people living there.

Keywords: Knowledge Discovery; data mining; multivariate statistical; correspondence analysis; digital terrain model; archaeology.

SUMÁRIO

LISTA DE FIGURAS	9
LISTA DE QUADROS	15
LISTA DE TABELAS	16
1 INTRODUÇÃO.....	17
1.1 Motivação	17
1.2 Objetivo	19
1.3 Contribuições.....	19
1.4 Organização do trabalho	19
2 O PROCESSO DA DESCOBERTA DE CONHECIMENTO.....	21
2.1 Visualização.....	23
3 MATERIAIS E MÉTODOS.....	26
3.1 Metodologia.....	26
3.2 Materiais	26
4 APLICAÇÃO DO PROCESSO DA DESCOBERTA DE CONHECIMENTO	29
4.1 Domínio da aplicação	29
4.1.1 Sítio Arqueológico Piracanjuba.....	29
4.2 Identificação do repositório de dados.....	32
4.3 Pré-processamento.....	32
4.3.1 Seleção dos atributos	33
4.3.2 Limpeza dos dados	34
4.3.3 Análise preliminar dos dados.	35
4.3.4 Visualização preliminar	35
4.3.5 Conclusões da análise preliminar dos dados e atributos selecionados.	43
4.3.6 Tratamento dos dados.....	46
4.4 Mineração de dados (data mining)	48
4.4.1 Tarefas de mineração de dados.....	48
4.4.2 Técnicas estatísticas.....	51
4.4.3 Agrupamento e redução de dimensionalidade.....	52
4.4.4 Análise de correspondência ou análise das associações.....	54

4.4.4.1 Inércia.....	60
4.4.4.2 Sugestões úteis para a interpretação dos resultados da análise de correspondência.....	64
4.5 A questão espacial.....	66
4.5.1 Conceitos básicos de Modelo Numérico de Terreno (MNT).....	67
4.6 Estratégias de mineração.....	71
4.6.1.1 Tabela de contingência fcomum.....	73
4.6.1.2 Tabela de contingência fraros.....	97
4.6.1.3 Tabela de contingência fraríssimos.....	113
4.6.1.4 Tabela de contingência excepcionais.....	128
4.7 Conclusões da mineração de dados.....	134
5 CONCLUSÃO.....	138
6 REFERÊNCIAS.....	140
APÊNDICE A – MÍDIA DIGITAL.....	144

LISTA DE FIGURAS

Figura 1 - O Processo da Descoberta de Conhecimento	22
Figura 2 - Padrão senoidal implícito da tabela apresentada no quadro 1.....	24
Figura 3 - Localização do Sítio Arqueológico Piracanjuba em imagem CBERS, bandas 2, 3, 4, realçada por componentes principais – 08/09/2004.	30
Figura 4 - Planta das escavações do Sítio Arqueológico Piracanjuba – Piraju, São Paulo.	31
Figura 5 - Espacialização e distribuição de frequência dos fragmentos cerâmicos.	36
Figura 6 - Atributos do material constituinte: (a) tipo de antiplástico, (b) material mineral, (c) caco moído, (d) carvão.....	37
Figura 7 - Atributo quanto à origem (parte do vaso) e forma do fragmento – classe.....	38
Figura 8 - Atributos quanto à fabricação: (a) espessura da parede, (b) queima, (c) dureza, (d) tratamento de superfície.	39
Figura 9 - Atributos quanto à fabricação: (a) decoração interna, (b) decoração externa, (c) técnica de manufatura.....	42
Figura 10 -Estado de conservação.....	43
Figura 11 -Histogramas da granulometria dos materiais constituintes: material mineral (a), caco moído (b).	47
Figura 12 -Níveis de informação do domínio da aplicação.....	49
Figura 13 -Agrupamento X Classificação	50
Figura 14 -Tabela de contingência.	54
Figura 15 -Regras geométricas auxiliares para a interpretação do gráfico de dispersão dos resultados da análise de correspondência.....	66
Figura 16 -Grade retangular regular	69
Figura 17 -Grade triangular	70
Figura 18 -Gráfico das dimensões derivadas – tabela fcomum.....	76

Figura 19 -Gráfico de dispersão das dimensões derivadas 1X2 das estacas e NSAs da tabela fcomum.....	77
Figura 20 -Gráfico de dispersão das dimensões derivadas 1X3 das estacas e NSAs da tabela fcomum.....	78
Figura 21 -Gráfico de dispersão das dimensões derivadas 2X3 das estacas e NSAs da tabela fcomum.....	78
Figura 22 -Gráfico de dispersão das dimensões derivadas 1X2 dos atributos_categorias da tabela fcomum	79
Figura 23 -Gráfico de dispersão das dimensões derivadas 1X3 dos atributos_categorias da tabela fcomum	80
Figura 24 -Gráfico de dispersão das dimensões derivadas 2X3 dos atributos_categorias da tabela fcomum	80
Figura 25 -Mapa temático do atributo espessura de parede categoria 2 (EPC_2).....	82
Figura 26 -Mapa temático do atributo espessura de parede categoria 3 (EPC_3).....	82
Figura 27 -Mapa temático do atributo queima categoria 3 (Q_3).....	83
Figura 28 -Mapa temático do atributo queima categoria 5 (Q_5).....	83
Figura 29 -Mapa temático do atributo queima categoria 6 (Q_6).....	84
Figura 30 -Gráfico de dispersão das dimensões derivadas 1X2 do NSA1 conjunto com os atributos_categorias da tabela fcomum.....	85
Figura 31 -Mapa temático do atributo caco moído categoria 0 (CMC_0) ou antiplástico categoria 1 (TP_1).....	86
Figura 32 -Mapa temático do atributo caco moído categoria 2 (CMC_2).	86
Figura 33 -Mapa temático do atributo caco moído categoria 3 (CMC_3).	87
Figura 34 -Mapa temático do atributo material mineral categoria 2 (MNC_2)	87
Figura 35 -Mapa temático do atributo material mineral categoria 3 (MNC_3)	88
Figura 36 -Mapa temático do atributo queima categoria 4 (Q_4).....	88
Figura 37 -Mapa temático do atributo caco moído categoria 1 (CMC_1).	89
Figura 38 -Mapa temático do atributo material mineral categoria 1 (MNC_1).	89
Figura 39 -Gráfico de dispersão das dimensões derivadas 2X3 das estacas, NSA2 e NSA3, conjunto com atributos_categorias da tabela fcomum,	

apresentando o correlacionamento positivo entre estacas e NSAs e atributos_categorias.	91
Figura 40 -Gráfico de dispersão das dimensões derivadas 2X3 das estacas, NSA2 e NSA3, conjunto com atributos_categorias da tabela fcomum, apresentando o correlacionamento negativo entre estacas e NSAs e atributos_categorias.	92
Figura 41 -Mapa temático do atributo espessura de parede categoria 1 (EPC_1).....	93
Figura 42 -Mapa temático do atributo decoração interna categoria 11 (DI_11).	93
Figura 43 -Mapa temático do atributo classe categoria 2 (CL_2).	94
Figura 44 -Mapa temático do atributo decoração interna categoria 15 (DI_15).	94
Figura 45 -Mapa temático do atributo decoração externa categoria 15 (DE_15).	95
Figura 46 -Mapa temático do atributo decoração externa categoria 11 (DE_11)	96
Figura 47 -Mapa temático do atributo espessura de parede categoria 4 (EPC_4).....	96
Figura 48 -Gráfico das dimensões derivadas – tabela fraros.....	99
Figura 49 -Gráfico de dispersão das dimensões derivadas 1X2 das estacas e NSAs da tabela fraros.....	100
Figura 50 -Gráfico de dispersão das dimensões derivadas 1X3 das estacas e NSAs da tabela fraros.....	100
Figura 51 -Gráfico de dispersão das dimensões derivadas 2X3 das estacas e NSAs da tabela fraros.....	101
Figura 52 -Gráfico de dispersão das dimensões derivadas 1X2 dos atributos_categorias da tabela fraros	102
Figura 53 -Gráfico de dispersão das dimensões derivadas 1X3 dos atributos_categorias da tabela fraros	102
Figura 54 -Gráfico de dispersão das dimensões derivadas 2X3 dos atributos_categorias da tabela fraros	103
Figura 55 -Mapa temático do atributo queima categoria 0 (Q_0).	103
Figura 56 -Mapa temático do atributo queima categoria 1 (Q_1).	104
Figura 57 -Gráfico de dispersão das dimensões derivadas 1X2 das estacas e NSAs conjunto com atributos_categorias da tabela fraros, apresentando a correlação positiva entre estacas e NSAs e atributos_categorias.	106

Figura 58 -Gráfico de dispersão das dimensões derivadas 1X2 das estacas e NSAs conjunto com atributos_categorias da tabela fraros, apresentando a correlação negativa entre estacas e NSAs e atributos_categorias.	107
Figura 59 -Mapa temático do atributo tratamento superficial categoria 5 (TS_5).	108
Figura 60 -Mapa temático do atributo tratamento superficial categoria 7 (TS_7).	108
Figura 61 -Mapa temático do atributo decoração externa categoria 10 (DE_10).	109
Figura 62 -Mapa temático do atributo decoração interna categoria 10 (DI_10).	109
Figura 63 -Mapa temático do atributo decoração interna categoria 16 (DI_16).	110
Figura 64 -Mapa temático do atributo decoração externa categoria 16 (DE_16).	110
Figura 65 -Mapa temático do atributo classe categoria 5 (CL_5).	111
Figura 66 -Mapa temático do atributo decoração externa categoria 4 (DE_4).	111
Figura 67 -Mapa temático do atributo queima categoria 2 (Q_2).	112
Figura 68 -Gráfico das dimensões derivadas – tabela fraríssimos	114
Figura 69 -Gráfico de dispersão das dimensões derivadas 1X2 das estacas e NSAs da tabela fraríssimos	115
Figura 70 -Gráfico de dispersão das dimensões derivadas 1X3 das estacas e NSAs da tabela fraríssimos	115
Figura 71 -Gráfico de dispersão das dimensões derivadas 2X3 das estacas e NSAs da tabela fraríssimos	116
Figura 72 -Gráfico de dispersão das dimensões derivadas 1X2X3 dos atributos_categorias da tabela fraríssimos	117
Figura 73 -Mapa temático do atributo decoração externa categoria 2 (DE_2).	117
Figura 74 -Mapa temático do atributo decoração externa categoria 5 (DE_5).	118
Figura 75 -Mapa temático do atributo tratamento superficial categoria 20 (TS_20)...	118
Figura 76 -Mapa temático do atributo decoração interna categoria 0 (DI_0).	119
Figura 77 -Mapa temático do atributo decoração externa categoria 0 (DE_0).	120
Figura 78 -Mapa temático do atributo decoração externa categoria 36 (DE_36)	120
Figura 79 -Mapa temático do atributo tratamento superficial categoria 0 (TS_0).	121

Figura 80 -Mapa temático do atributo tratamento superficial categoria 11 (TS_11)...	121
Figura 81 -Gráfico de dispersão das dimensões derivadas 1X2 das estacas e NSAs conjunto com atributos_categorias da tabela fraríssimos, apresentando a correlação positiva entre estacas e NSAs e atributos_categorias.	122
Figura 82 -Gráfico de dispersão das dimensões derivadas 1X2 das estacas e NSAs conjunto com atributos_categorias da tabela fraríssimos, apresentando a correlação negativa entre estacas e NSAs e atributos_categorias.	123
Figura 83 -Mapa temático do atributo técnica de manufatura categoria 3 (MAN_3)..	124
Figura 84 -Mapa temático do atributo classe categoria 3 (CL_3).....	124
Figura 85 -Mapa temático do atributo classe categoria 23 (CL_23).....	125
Figura 86 -Mapa temático do atributo decoração externa categoria 14 (DE_14).	125
Figura 87 -Mapa temático do atributo decoração interna categoria 14 (DI_14).....	126
Figura 88 -Mapa temático do atributo tratamento superficial categoria 6 (TS_6).	126
Figura 89 -Mapa temático do atributo tratamento superficial categoria 12 (TS_12)...	127
Figura 90 -Mapa temático do atributo tratamento superficial categoria 19 (TS_19)...	127
Figura 91 -Gráfico de colunas dos fragmentos cerâmicos coletados no NSA1 que apresentam o atributo classe em suas categorias excepcionais.	129
Figura 92 -Gráfico de colunas dos fragmentos cerâmicos coletados no NSA1 que apresentam o atributo tratamento superficial em suas categorias excepcionais.....	129
Figura 93 -Gráfico de colunas dos fragmentos cerâmicos coletados no NSA1 que apresentam o atributo decoração interna em suas categorias excepcionais.....	130
Figura 94 -Gráfico de colunas dos fragmentos cerâmicos coletados no NSA1 que apresentam o atributo decoração externa em suas categorias excepcionais.....	130
Figura 95 -Gráfico de colunas dos fragmentos cerâmicos de ocorrência exclusiva no NSA1 de atributos_categorias excepcionais.....	131
Figura 96 -Gráfico de colunas dos fragmentos cerâmicos coletados no Sítio Piracanjuba, exceto o NSA1, que apresentam o atributo classe em suas categorias excepcionais.	132
Figura 97 -Gráfico de colunas dos fragmentos cerâmicos coletados no Sítio Piracanjuba, exceto o NSA1, que apresentam o atributo decoração	

interna em suas categorias excepcionais.133

Figura 98 -Gráfico de colunas dos fragmentos cerâmicos coletados no Sítio Piracanjuba, exceto o NSA1, que apresentam o atributo decoração externa em suas categorias excepcionais.....133

Figura 99 -Mapa temático de áreas diferenciadas quanto aos atributos_categorias do Sítio Piracanjuba.....137

LISTA DE QUADROS

Quadro 1 - Tabela com padrão de difícil percepção	24
Quadro 3 - Atributos registrados quanto à proveniência.	27
Quadro 4 - Atributos registrados quanto ao material constituinte.	27
Quadro 5 - Atributos registrados quanto à fabricação.	27
Quadro 6 - Atributos registrados quanto à origem (parte do vaso) e forma do fragmento.....	27
Quadro 7 - Atributos registrados quanto ao uso e conservação.....	27
Quadro 8 - Atributos selecionados para o processo da descoberta de conhecimento.....	44
Quadro 9 - Descrição dos atributos_categorias selecionados para análise.	44
Quadro 10 - Categorização dos atributos material mineral, caco moído, carvão e espessura da parede.....	47
Quadro 11 - Tarefas de mineração de dados e as técnicas associadas a elas.....	49
Quadro 12 - Intervalos de fatiamento para a ocorrência de excesso/falta de atributos_categorias de fragmentos cerâmicos	70
Quadro 13 - Características das tabelas para análise.	71

LISTA DE TABELAS

Tabela 1 - Valores singulares, autovalores e inércia – tabela fcomum.....	75
Tabela 2 - Valores singulares, autovalores e inércia – tabela fraros.....	98
Tabela 3 - Valores singulares, autovalores e inércia – tabela fraríssimos.....	114

INTRODUÇÃO

1.1 MOTIVAÇÃO

Durante os últimos anos tem se verificado um crescimento substancial na quantidade de dados armazenados em meios magnéticos ou não em todas as áreas da atuação humana. Estes dados, produzidos e armazenados em larga escala, são inviáveis de serem lidos ou analisados por especialistas utilizando métodos manuais tradicionais (PIATETSKY-SHAPIRO, 1991), tais como planilhas de cálculos e relatórios informativos operacionais, nos quais o especialista testa sua hipótese contra a base de dados. Por outro lado, sabe-se que grandes quantidades de dados equivalem a um maior potencial de informação. Entretanto, as informações contidas nos dados não estão caracterizadas explicitamente, uma vez sendo dados brutos e não interessando quando estudados individualmente. Diante deste cenário, surge a necessidade de se explorar estes dados para extrair informação - conhecimento explícito, e utilizá-lo no âmbito do problema.

Essa conjuntura levou ao surgimento de uma área de pesquisa conhecida como Descoberta de Conhecimento em Banco de Dados - “*Knowledge Discovery Database*” (*KDD*), tendo sido definida por Fayyad et al. (1996) como “Um processo não trivial de identificação de padrões válidos, novos, úteis e implicitamente presentes em grandes volumes de dados”, cujo cerne é composto pelas técnicas de mineração de dados – “*Data Mining*” (*DM*).

“O arqueólogo não anda à procura de “antiguidades”, mas de conhecimentos, de realizações inteligentes” (MOBERG, 1986).

Artefatos cerâmicos, dentro da arqueologia, são considerados como elementos fundamentais na obtenção de informações e no entendimento de como diferentes civilizações

viviam (MUNITA et al., 2001). Esse fato aliado a descobertas de sítios arqueológicos, segundo o Instituto do Patrimônio Histórico e Artístico Nacional (IPHAN) o Brasil possui atualmente 20.000 sítios arqueológicos, alguns com grande número de artefatos cerâmicos (milhares), propiciam excelente arsenal na investigação e conhecimento de civilizações que outrora habitaram o País.

Existem estudos versando sobre a análise de artefatos cerâmicos por meio de técnicas de estatística multivariada com enfoque em sua composição mineralógica (BAXTER et al., 2003; MUNITA et al., 2001, 2004), bem como aqueles que utilizando-se da estatística descritiva realizam um levantamento considerando aspectos morfológicos, topológicos e decorativos (FACCIO, 1992). Todavia, a análise de artefatos cerâmicos considerando aspectos morfológicos, topológicos, decorativos e espaciais por meio de técnicas de mineração de dados é incipiente.

Neste contexto, este trabalho apresenta os resultados obtidos em um processo de descoberta de conhecimento no Sítio Arqueológico Piracanjuba, localizado no Município de Pirajú – SP, empregando, para tanto, atributos e técnicas capazes de extrair do conjunto de dados informações passíveis de agregar conhecimento ao estudo de antigas civilizações, contribuindo desta forma para os estudos iniciados em 1968, com a criação do Projeto Paranapanema (ProjPar).

As técnicas estatísticas de mineração de dados aplicadas ao conjunto de atributos dos artefatos cerâmicos coligidos ao nível do solo objetivam reduzir a dimensionalidade dos dados e promover a associação dos atributos morfológicos, topológicos e decorativos.

Os resultados obtidos na mineração de dados são apresentados sob a forma de gráficos e mapas temáticos, que permitem a análise visual do comportamento espacial dos atributos, propiciando aos arqueólogos pesquisadores uma visão sistêmica do Sítio Arqueológico Piracanjuba.

1.2 OBJETIVO

Prover os arqueólogos pesquisadores de uma visão sistêmica do Sítio Piracanjuba capaz de agregar conhecimento ao estudo das populações pretéritas que ali habitavam.

Os objetivos específicos são:

- definir as técnicas capazes de extrair conhecimento do conjunto de artefatos cerâmicos coletados no nível zero, respeitando as restrições impostas pelo conjunto de dados;
- apresentar os resultados de maneira a permitir a extração visual de conhecimento por peritos da arqueologia;
- integrar as necessidades da arqueologia de forma inovadora a outras áreas do conhecimento humano como informática, estatística e análise espacial.

1.3 CONTRIBUIÇÕES

Este trabalho deverá propiciar maior conhecimento sobre os antigos habitantes do Sítio Arqueológico Piracanjuba e lançar as bases para o desenvolvimento de um sistema de descoberta de conhecimento que se adéque às necessidades dos arqueólogos pesquisadores.

1.4 ORGANIZAÇÃO DO TRABALHO

O capítulo presente tem como fundamento situar e definir os objetivos do trabalho e tem continuidade nos capítulos a seguir.

Considerando a multidisciplinaridade envolvida neste estudo, os conceitos e as técnicas concernentes às diversas áreas envolvidas na consecução deste trabalho foram apresentados em capítulos distintos.

O capítulo 2 englobou conceitos da descoberta de conhecimento em banco de dados

e visualização.

O capítulo 3 apresenta a metodologia e os materiais selecionados para o desenvolvimento do estudo

O capítulo 4 apresenta:

- o domínio da aplicação;
- o pré-processamento dos dados;
- a mineração de dados, suas estratégias, aplicação, os resultados obtidos e conclusões;
- o tratamento dado à questão espacial.

A conclusão, capítulo 5, apresenta uma análise quanto à empregabilidade da metodologia e das técnicas selecionadas, tendo por base os resultados obtidos neste trabalho.

O PROCESSO DA DESCOBERTA DE CONHECIMENTO

O processo capaz de descobrir conhecimento em banco de dados chama-se KDD (Knowledge Discovery Database) tendo sido proposto em 1989 para referir-se às etapas que produzem conhecimento a partir dos dados e, principalmente, à etapa de mineração dos dados, que é a fase que transforma dados em informações.

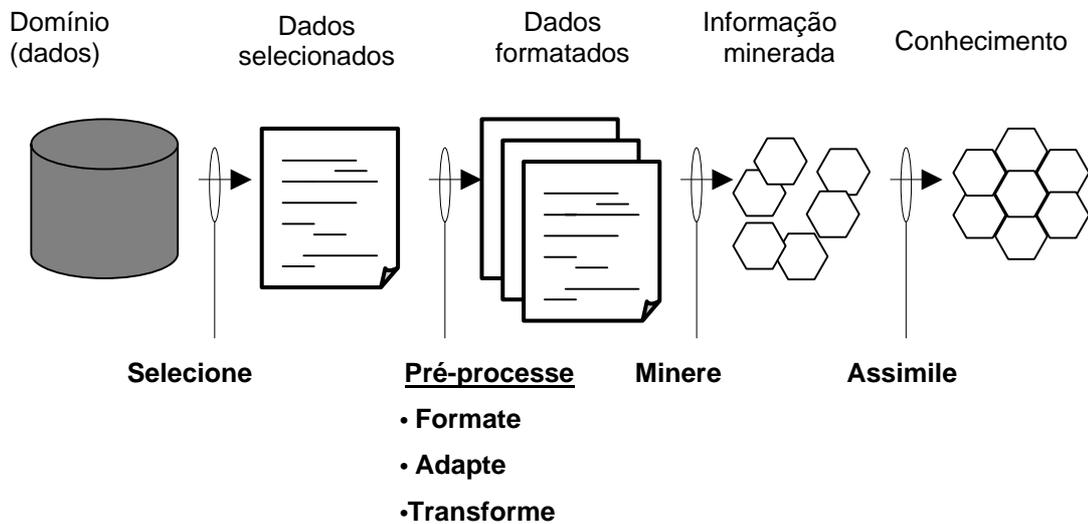
A descoberta de conhecimento pode ser entendida como uma mudança de filosofia entre uma abordagem de análise de dados voltada à verificação, para uma voltada à descoberta (MENDONÇA NETO, 2001). Em uma abordagem voltada à descoberta, os dados coletados são “filtrados” para a extração de informações capazes de agregar conhecimento novo, potencialmente útil, previamente desconhecido e não trivial a um determinado domínio de aplicação, entendendo-se como domínio da aplicação o sistema real ou abstrato sob análise por meio dos dados coletados.

Um processo típico de descoberta de conhecimento possui quatro passos: seleção de dados, pré-processamento dos dados, extração de informações e a assimilação da informação minerada, conforme ilustra a figura 1.

Dados brutos coletados de um ou mais repositório de dados dificilmente se encontram prontos para a mineração, devendo ser preparados para tal, o que envolve as fases de pré-processamento e tratamento dos dados. Algumas operações comuns desta fase são: escolha dos dados relevantes, padronização, eliminação de redundância, simplificação da estrutura e a limpeza de dados de maneira a se obter uma base representativa no formato atributo/valor (BIGOLIN, 2000).

A mineração de dados é o passo em que são aplicadas técnicas voltadas para atingir objetivos específicos dentro da área de atuação humana da aplicação (GROEBEL e GRUENWALD, 1999). A escolha das técnicas de mineração de dados é função da tarefa específica a ser executada e dos dados disponíveis para a análise, sendo a familiaridade com

as técnicas indispensável para a melhor abordagem do problema apresentado (FAQ, 1996).



Fonte: MENDONÇA NETO (2001).

FIGURA 1 - O Processo da Descoberta de Conhecimento

Segundo Cratochvil (1999), a mineração de dados combina métodos e técnicas de diversas áreas do conhecimento humano, sendo a aprendizagem de máquina, estatística, banco de dados, sistemas especialistas e visualização de dados as mais relacionadas.

Desta forma, em um processo completo de descoberta de conhecimento podem ser utilizadas diversas técnicas de mineração de dados, especificadas em função do nível de conhecimento a ser extraído e da área de conhecimento humano da aplicação. As técnicas dão suporte a um conjunto de tarefas: associações, padrões seqüenciais, séries temporais similares, classificação, regressão e agrupamentos, que diferem entre si pelo tipo de problema que são capazes de resolver (IBM, 1996).

O último passo da descoberta do conhecimento envolve a interpretação e assimilação da informação minerada expressa por padrões ou modelos.

1.5 VISUALIZAÇÃO

Dentre as formas que o ser humano utiliza para interagir com o meio, seja para reconhecimento ou aprendizado, a mais eficiente é a visão humana. Para ter uma idéia do potencial representado pela visão humana, segundo Ferreira (1975), a porcentagem de retenção mnemônica do ser humano é a seguinte: 1% por meio do gosto, 1,5% por meio do tato, 3,5% por meio do olfato, 11% por meio da audição e 83% por meio da visão.

Os seres humanos usam meios visuais não só para extrair, associar ou expandir conhecimento, usam também meios gráficos para representar informações para outros seres humanos e como facilitadores na análise de sistemas abstratos ou não, nos quais a extração do conhecimento seria inviável de outra forma, seja pelo número de fatores envolvidos ou pela complexidade dos mesmos.

Dentro deste contexto, o uso dos computadores é útil para a representação visual de dados abstratos, pois permitem aglutinar centenas ou até milhares de dados em uma única imagem, criando um meio eficiente na interpretação de grande volume de dados. Por outro lado, a preocupação com a forma pela qual a informação será exibida possui aspecto relevante, pois ela deve ser clara, não ambígua e permitir rápida assimilação pelo ser humano.

Para clarificar os conceitos até aqui introduzidos será apresentado, a seguir, um exemplo retirado de Mendonça (2001).

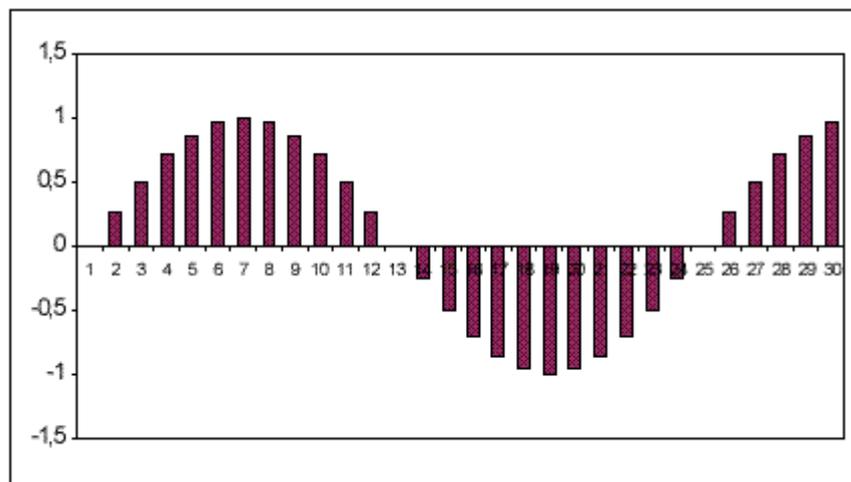
Observe os dados do quadro 1 e veja que o padrão nele apresentado é de difícil memorização e interpretação por seres humanos.

Porém, o padrão periódico senoidal da tabela apresentada no quadro 1 é facilmente interpretado e processado pelo cérebro humano quando apresentado no formato gráfico ilustrado na figura 2.

Quadro 1 - Tabela com padrão de difícil percepção

X	0	0,259	0,5	0,707	0,866	0,966	1	0,966	0,866	0,707
Y	12	13	14	15	16	17	18	19	20	21
X	0,5	0,259	0	-0,259	-0,5	-0,707	-0,866	-0,966	-1	-0,966
Y	22	23	24	25	26	27	28	29	30	31
X	-0,866	-0,707	-0,5	-0,259	0	0,259	0,5	0,707	0,866	0,966
Y	32	33	34	35	36	37	38	39	40	41

Fonte: MENDONÇA NETO (2001).



Fonte: MENDONÇA NETO (2001).

FIGURA 2 - Padrão senoidal implícito da tabela apresentada no quadro 1.

A capacidade e a facilidade de processar dados visuais do ser humano fizeram surgir progressivas áreas de pesquisa como a que trata da interação ser humano – representação de dados e informações por meios gráficos conhecida como Visualização, interação ser humano – máquina por meios gráficos conhecida como Interface homem – máquina.

A visualização está presente em todos os passos do processo da descoberta de conhecimento desde a seleção de atributos, útil no estudo da distribuição dos dados, na detecção de dados omissos ou inconsistentes, até a fase de apresentação da informação minerada ao especialista. O uso de métodos adequados de visualização deve dar condições de interpretar e extrair conhecimento a partir de informações complexas, ou seja, boas técnicas

de visualização devem auxiliar o cérebro humano a processar informações complexas de forma clara, rápida e eficiente.

Convém ressaltar que a apresentação visual dos resultados obtidos pelas técnicas de mineração de dados possibilita a sua melhor interpretação e assimilação, posto que padrões minerados geralmente são complexos e variam em muitas dimensões, o que dificulta a sua interpretação quando apresentados na forma de tabelas ou textos.

MATERIAIS E MÉTODOS

1.6 METODOLOGIA

A metodologia proposta encontra-se descrita em Berry e Linoff (1997) e é conhecida como descoberta de conhecimento indireto. Os passos que a integram são descritos sumariamente abaixo:

- identificar os repositórios de dados disponíveis;
- preparar os dados para análise;
- selecionar as técnicas apropriadas com base nas características dos dados e na tarefa de mineração de dados;
- usar os resultados para descobrir estruturas escondidas nos dados;
- identificar alvos em potencial para a descoberta de conhecimento indireto;
- gerar hipóteses a serem testadas.

Cabe ressaltar o fato da descoberta de conhecimento ser um processo iterativo e iterativo, portanto todas as fases metodológicas são passíveis de reavaliação, repetição e modificação.

1.7 MATERIAIS

- Dados arqueológicos do Sítio Piracanjuba, Piraju-SP, contendo 15.288 registros de fragmentos cerâmicos em planilha eletrônica, com atributos registrados em grupos quanto à natureza das informações, ilustrados a seguir: quanto à identificação, quadro 2; quanto à proveniência, quadro 3; quanto ao material constituinte, quadro 4; quanto à fabricação, quadro 5; quanto à origem (parte do vaso) e forma do fragmento, quadro 6; e quanto ao uso e conservação, quadro 7.

Quadro 2 - Atributos registrados quanto à identificação.

Atributo	Número da Peça	número do Vaso
Abrev.	PEÇA	VASO
Tipo	nomi.	nomi.

Quadro 3 - Atributos registrados quanto à proveniência.

Atributo	Setor	Estaca	Leiria	Qua- dra	Quadrí- cula	Trin- cheira	Corte de Verificação	Área de Decapagem	Perfil	Nível
Abrev.	Setor	Estaca	Leiria	QDA	QDI	T	CV	D	P	NI
Tipo	nomi.	nomi.	nomi.	nomi.	nomi.	nomi.	nomi.	nomi.	nomi.	nomi.

Quadro 4 - Atributos registrados quanto ao material constituinte.

Atributo	Tipo do Antiplástico	Material mineral (mm)	Cariapé (mm)	CacoMoído (mm)	Carvão (mm)	Concha Moída (mm)
Abrev.	TP	MN	CR	CM	CV	CC
Tipo	nomi.	ordinal	ordinal	ordinal	ordinal	ordinal

Quadro 5 - Atributos registrados quanto à fabricação.

Atributo	Espessura da parede (mm)	Queima	Dureza	Cor da Argila	Tratamento Superficial	Deco- ração Interna	Deco- ração Externa	Técnica de manufatura
Abrev.	ES.PAR	QUEIMA	DUREZ	COR.ARG	TRA.SUP.	DECORI	DECORE	TEC.MAN.
Tipo	ordinal	nomi.	nomi.	nomi.	nomi.	nomi.	nomi.	nomi.

Quadro 6 - Atributos registrados quanto à origem (parte do vaso) e forma do fragmento.

Atributo	Classe	Ângulo da Parede	Bolota de Argila	Forma de Lábio	Forma da Borda	Forma da Vasilha	Contorno do recipiente
Abrev.	CLASSE	ANG. PAR	BOLOT. ARG.	LÁBIO	BORDA.	FORMA	CONTOR
Tipo	nomi.	nomi.	nomi.	nomi.	nomi.	nomi.	nomi.

Atributo	Diâmetro da Boca (mm)	Altura do Vaso (mm)	Largura da Garganta (mm)	Volume do Vaso (cm ³)	Tipo de Base	Diâmetro da Base (mm)	Ângulo da Base (graus)
Abrev.	D. BOCA.	ALT. VASO	LAR. GAR	V. VASO	BASE	D. BASE	AG. BASE
Tipo	ordinal	ordinal	ordinal.	ordinal	nomi.	ordinal	ordinal

Quadro 7 - Atributos registrados quanto ao uso e conservação.

Atributo	Marcas de Uso	Estado de Conservação
Abrev.	MAR.USO	EST.CONS.
Tipo	nomi.	nomi.

- Carta topográfica do sítio com localização do berço dos artefatos cerâmicos.
- Software estatístico para processamento, geração de resultados e gráficos.
- Software de análise espacial para a geração de modelos numéricos de terreno e de mapas temáticos.

APLICAÇÃO DO PROCESSO DA DESCOBERTA DE CONHECIMENTO

1.8 DOMÍNIO DA APLICAÇÃO

Segundo Mendonça Neto (2001), entende-se como domínio da aplicação o sistema real ou abstrato sob análise por meio dos dados coletados, em que dado é o conjunto de símbolos usados para se representar um valor associado a certo atributo de uma certa entidade, sendo uma entidade um membro distinto desse domínio caracterizada por atributos e relações com outras entidades. Portanto, para fins do estudo o domínio da aplicação é o Sítio Arqueológico Piracanjuba analisado pelos dados dos fragmentos cerâmicos ali coletados e caracterizados por seus atributos de proveniência, material constituinte, fabricação, origem (parte do vaso) e forma do fragmento, uso e conservação.

1.8.1 Sítio Arqueológico Piracanjuba

O Sítio Arqueológico Piracanjuba está localizado no Município de Piraju, Estado de São Paulo, no médio curso do Rio Paranapanema com coordenadas UTM 656.750 E e 7.438.600 N, conforme ilustra a figura 3.

Para a área desse sítio foram realizados levantamentos planialtimétricos, sondagens, perfis, trincheiras e decapagens.

O sítio pode ser classificado como razoavelmente conservado, as alterações na superfície do terreno chegam a ser significativas, comprometendo parte do registro arqueológico. Os estratos arqueológicos chegam a ser afetados por sulcos de erosão, bioturbação e pela ação de implementos agrícolas ou outras atividades antrópicas.

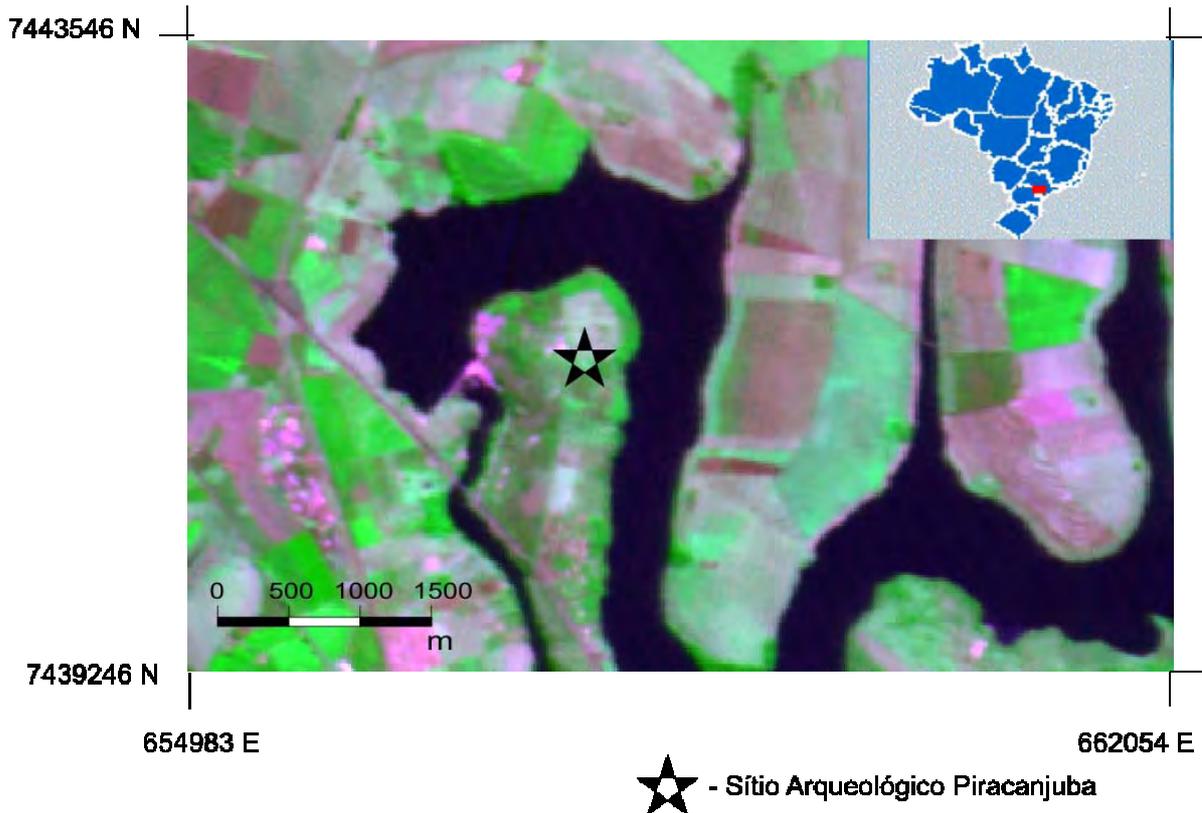


FIGURA 3 - Localização do Sítio Arqueológico Piracanjuba em imagem CBERS, bandas 2, 3, 4, realçada por componentes principais – 08/09/2004.

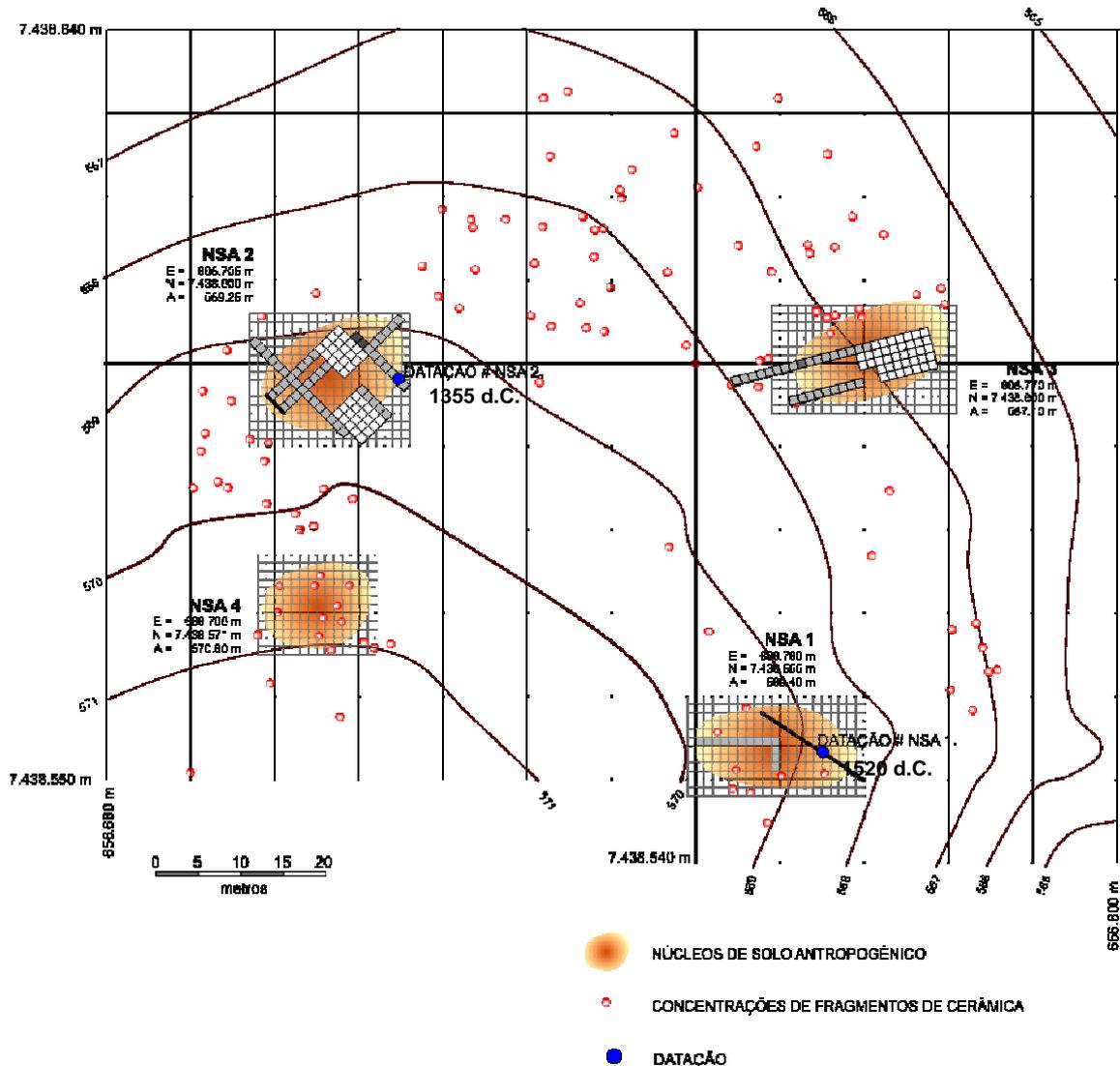
O trabalho de pesquisa nesse sítio foi iniciado no ano 2000 e faz parte do Programa de Salvamento Arqueológico da UHE Piraju, coordenado pelo Prof. Dr. José Luiz de Moraes do Museu de Arqueologia e Etnologia MAE / Universidade de São Paulo (USP).

O material arqueológico aflorou por ocasião de escavações realizadas na abertura de valas para depósito de lixo pela Companhia Brasileira de Alumínio (CBA), quando o coordenador do projeto tomou as medidas necessárias junto ao IPHAN e à referida empresa, para preservar o patrimônio arqueológico contido na área do sítio. Após o cadastro do sítio no IPHAN e da autorização do referido órgão, os trabalhos de campo e de laboratório tiveram início.

O Sítio Piracanjuba tornou-se um Sítio Escola pelo trabalho de uma equipe interdisciplinar, no qual alunos do curso de pós-graduação em arqueologia do MAE/USP, bem como alunos de graduação dos Cursos de Geografia e História da USP, anualmente têm a

oportunidade de conhecer na prática os métodos do trabalho de campo.

SÍTIO ARQUEOLÓGICO PIRACANJUBA PROJETO PARANAPANEMA



Fonte: Projeto Paranapanema – Centro Regional de Arqueologia Ambiental em Piraju Contrato Universidade de São Paulo – Companhia Brasileira de Alumínio (2004).

FIGURA 4 - Planta das escavações do Sítio Arqueológico Piracanjuba – Piraju, São Paulo.

Até o momento foram coletados sistematicamente 15.288 fragmentos cerâmicos em 4 fundos de habitações e entornos: Núcleo de Solo Antropogênico 1 (NSA1), Núcleo de Solo Antropogênico 2 (NSA2), Núcleo de Solo Antropogênico 3 (NSA3), Núcleo de Solo Antropogênico 4 (NSA4), além de coleta de superfície marcada por estacas e leirias, em

concentrações de fragmentos de cerâmica, ilustrados na figura 4.

A datação do NSA1 realizada em 2000 e a do NSA2 em 2004 apontam para os anos de 1520 d.C. e 1355 d.C., respectivamente.

Segundo Moraes e Faccio (2004), o material cerâmico recuperado pertence, de modo geral, a denominada Tradição Tupiguarani, encontrando-se devidamente limpo, registrado e acervado no Centro de Pesquisas Mario Neme da Universidade de São Paulo.

Estudos que contemplam a vasilha enquanto objeto de análise, bem como a análise dos fragmentos por meio de técnicas da estatística descritiva constam do trabalho intitulado “O trabalho do barro no Sítio Arqueológico Piracanjuba – Piraju, SP.” (MORAIS e FACCIO, 2004).

1.9 IDENTIFICAÇÃO DO REPOSITÓRIO DE DADOS.

Sendo uma das bases da descoberta do conhecimento a premissa de que dados necessários à mineração de dados em determinado domínio estão disponíveis em algum repositório de dados, o analista deve ter garantias quanto à integridade e à confiabilidade dos mesmos, realizando então uma pré-seleção dos dados a serem utilizados pelo algoritmo de mineração.

O repositório de dados foi fornecido pela equipe de arqueólogos do ProjPar, em planilha eletrônica contendo os registros de fragmentos cerâmicos coletados no Sítio Piracanjuba – Piraju, SP.

1.10 PRÉ-PROCESSAMENTO

Foram mapeados na área do sítio, até o momento, quatro Núcleos de Solo Antropogênico, além de uma grande quantidade de material arqueológico em superfície no

entorno desses núcleos. Dos quatro Núcleos de Solo Antropogênico identificados na área do sítio, três já tiveram algum tipo de intervenção. Contudo, as escavações desses níveis estão apenas no início.

Para a presente pesquisa foram selecionados os registros de fragmentos cerâmicos coletados ao nível do solo em 3 Núcleos de Solo Antropogênico, NSA1, NSA2 e NSA3, e 100 estacas dispersas em uma gleba de terreno de aproximadamente 12.000 m², perfazendo então um total de 7564 registros de fragmentos cerâmicos para análise.

1.10.1 Seleção dos atributos

A escolha dos atributos relevantes é uma das fases mais importantes no processo da descoberta do conhecimento, estando toda a qualidade intimamente ligada a ela, valendo aqui a máxima da computação “entra lixo, sai lixo”, pois dela vai depender a escolha do algoritmo e o nível de mineração a ser empregado na fase da mineração dos dados.

Essa escolha teve por base a metodologia desenvolvida por Robrahn (1996), posteriormente adaptada por Faccio (1998) em atendimento às especificidades da área do ProjPar, e à consistência dos dados registrados para cada atributo.

Para Moraes e Faccio (2004), “... esta metodologia permite caracterizar e oferecer elementos para discussão e compreensão da seqüência dos gestos técnicos realizados para a produção dos artefatos da indústria cerâmica (cadeia operatória)”, tendo sido utilizado em materiais da área do ProjPar objetivando a curto prazo a possibilidade de comparar os materiais dos vários sítios da área do projeto e assim caracterizar o Sistema Regional Guarani instalado no Vale do Rio Paranapanema (LOPES e FACCIO, 2001).

Segundo Faccio (1998), “A análise do material cerâmico compreende a verificação

de classes de atributos tecnológicos, estilísticos e morfológicos, além das marcas de uso e do estado de conservação”. Portanto, no sentido de integrar este estudo aos realizados na área do ProjPar, foram considerados atributos relevantes e suas classes os definidos por Faccio (1998), aos quais foram incorporados atributos espaciais, coordenadas UTM dos berços dos fragmentos cerâmicos, de forma a permitir a análise da distribuição espacial desses atributos e suas classes de forma isolada e conjunta.

1.10.2 Limpeza dos dados

Na fase de limpeza dos dados foram eliminados os atributos sem dados registrados, dados não válidos, bem como aqueles que apresentavam mais de 90% de dados omissos, sendo eles:

- quanto à procedência: Leiria (Leiria), Quadra (QDA), Quadrícula (QDI), Trincheira (T), Corte de verificação (CV), área de decapagem (D) e perfil (P);
- quanto ao material constituinte: tamanho das partículas de cariapé (CR) e concha moída (CC);
- quanto à fabricação: cor da argila (C.ARG.);
- quanto à origem (parte do vaso) e forma do fragmento: Bolota de argila (BOLOT.ARG), contorno do recipiente (CONTOR.) e largura da garganta (LAR. GAR), e ângulo da base (AG. BASE).

1.10.3 Análise preliminar dos dados.

Antes de escolher a técnica e o nível de mineração de dados é necessária a análise e compreensão dos mesmos. Para tanto a estatística descritiva (média, moda, desvio padrão, distribuição dos dados) e as ferramentas de visualização, entendendo visualização como o uso de computadores para representação gráfica de dados (SANTOS et al., 2000), constituem-se em poderosas armas, pois possibilitam conhecimento “*a priori*” do conjunto de dados.

A análise preliminar dos dados enfocou critérios como: o número, a validade e a distribuição das observações realizadas para cada atributo considerado válido para este estudo.

Durante a execução deste trabalho foram realizadas várias reuniões com os orientadores no sentido de dirimir dúvidas e procurar caminhos. As reuniões com a orientadora Prof^a. Dr^a. Neide Faccio Barocá versaram sobre assuntos relacionados com o domínio da aplicação, tendo fundamental importância na fase da análise preliminar, parte delas compoendo este trabalho como: Palavra da especialista.

1.10.4 Visualização preliminar

Dentre as ferramentas de visualização bastante utilizadas encontram-se os gráficos de coluna, os de distribuição de frequência (histogramas) e os de dispersão em 2 ou 3 dimensões de diferentes atributos que mostram claramente os valores omissos, os valores excepcionais (“*outliers*”) e os relacionamentos dos atributos sob análise.

A análise preliminar foi elaborada tendo por base gráficos de coluna e objetiva

estudar a distribuição, a variabilidade e determinar o senso comum (moda) dos atributos.

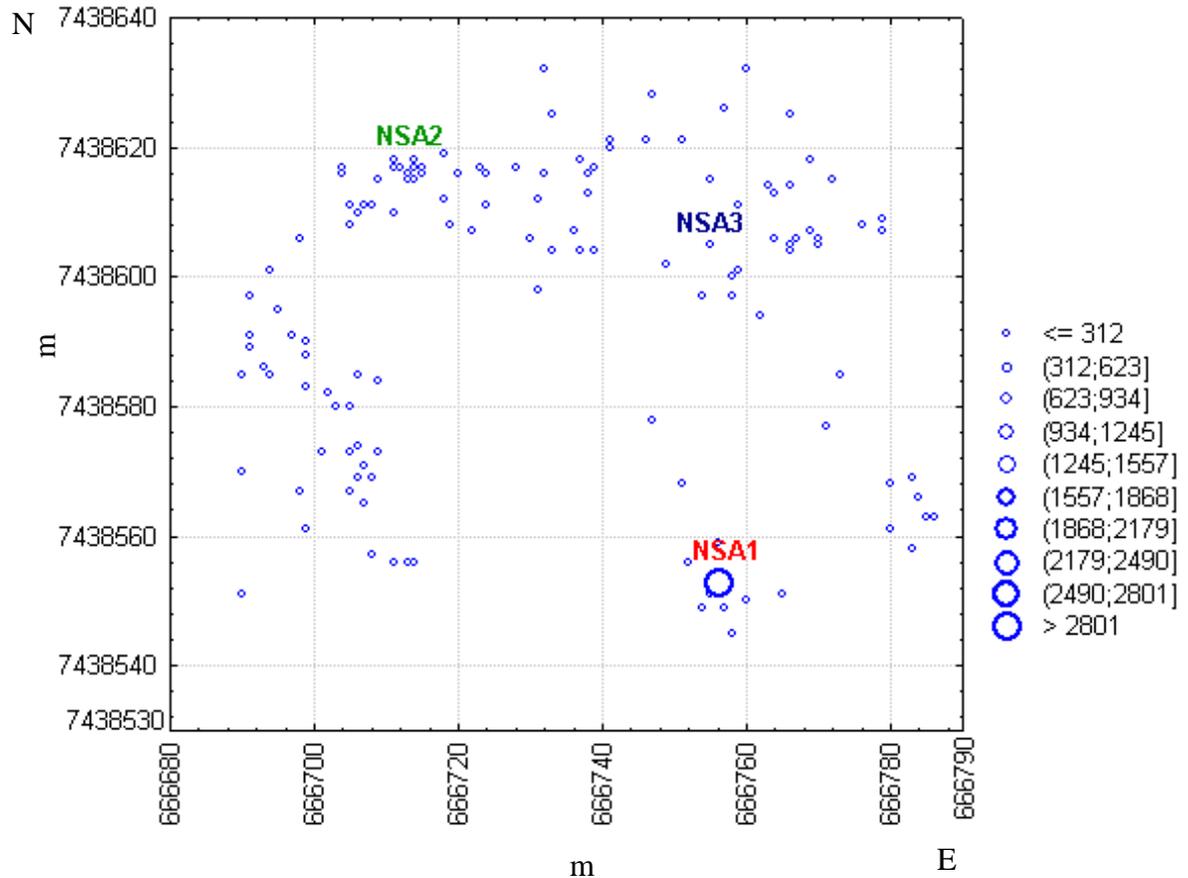
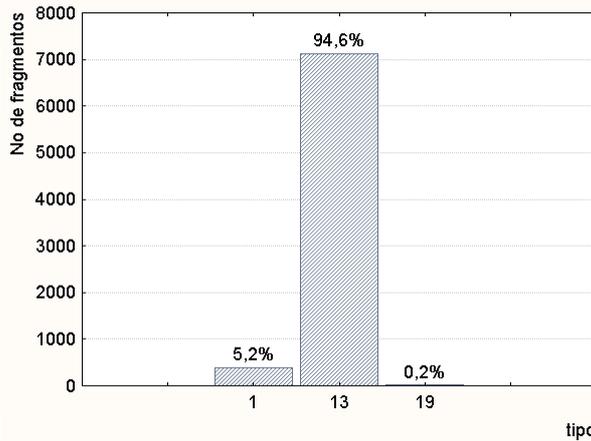


FIGURA 5 - Espacialização e distribuição de frequência dos fragmentos cerâmicos.

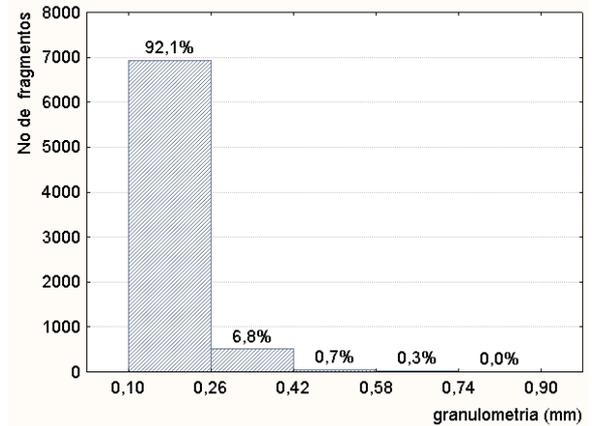
Quanto à procedência ilustrada na figura 5, gráfico de distribuição de frequências em duas dimensões (coordenadas UTM), pode-se dizer que existe concentração de fragmentos nos NSA1 e NSA2, sendo que o NSA1 contribui com o maior número de fragmentos coletados, acima de 2801, ocorrendo a maior dispersão no entorno do NSA3.

Quanto ao material constituinte ilustrado por seus atributos na figura 6, nos gráficos:

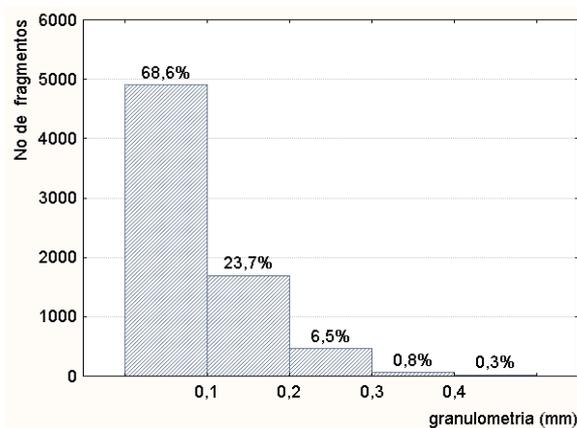
(a) antiplástico, (b) material mineral, (c) caco moído e (d) carvão, pode-se dizer que:



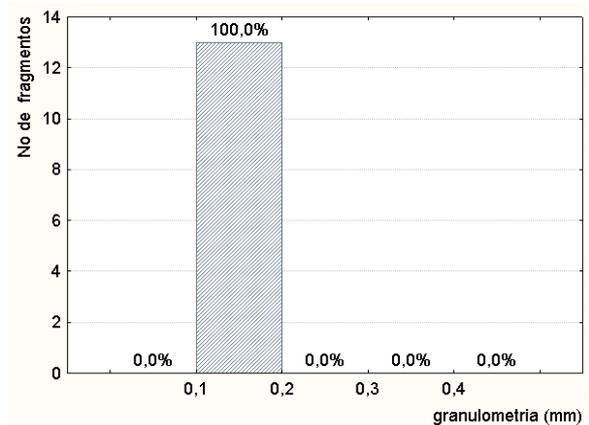
(a) Antiplástico



(b) Material Mineral



(c) Caco Moído



(d) Carvão

FIGURA 6 - Atributos do material constituinte: (a) tipo de antiplástico, (b) material mineral, (c) caco moído, (d) carvão.

- para o antiplástico, 94,6% dos fragmentos pertencem à categoria 13 (material mineral + caco moído), 5,2% à categoria 1 (material mineral) e 0,2%, à categoria 19 (material mineral + caco moído + carvão);
- para a granulometria do material mineral, 92,1% dos fragmentos são constituídos de material mineral variando entre 0,10 e 0,26 mm, 6,8% entre 0,26 e 0,42 mm, 0,7% entre 0,42 e 0,58 mm. e 0,3% acima de 0,58 mm;
- para o caco moído, 68,6% dos fragmentos apresentam granulometria menor que 0,1 mm, 23,3% entre 0,1 e 0,2 mm, 6,5% entre 0,2 e 0,3 mm e 0,3% acima de 0,4 mm, e
- para a granulometria do carvão, 100% dos fragmentos variam de 0,1 a 0,2 mm.

Quanto à origem (parte do vaso) e forma do fragmento – classe - ilustrada na figura 7, o gráfico de colunas do atributo classe mostra que: 82,0% dos fragmentos pertencem à categoria 1 (parede), 14,4% à 2 (borda), 0,4% à 3 (base), 0,1% à 4 (base, parede, borda), 1,8% à 5 (parede angular), 0,1% à 19 (parede com furo de suspensão), 0,3 % à 21 (borda com suporte para tampa), 0,9% à 23 (borda/ parede angular).

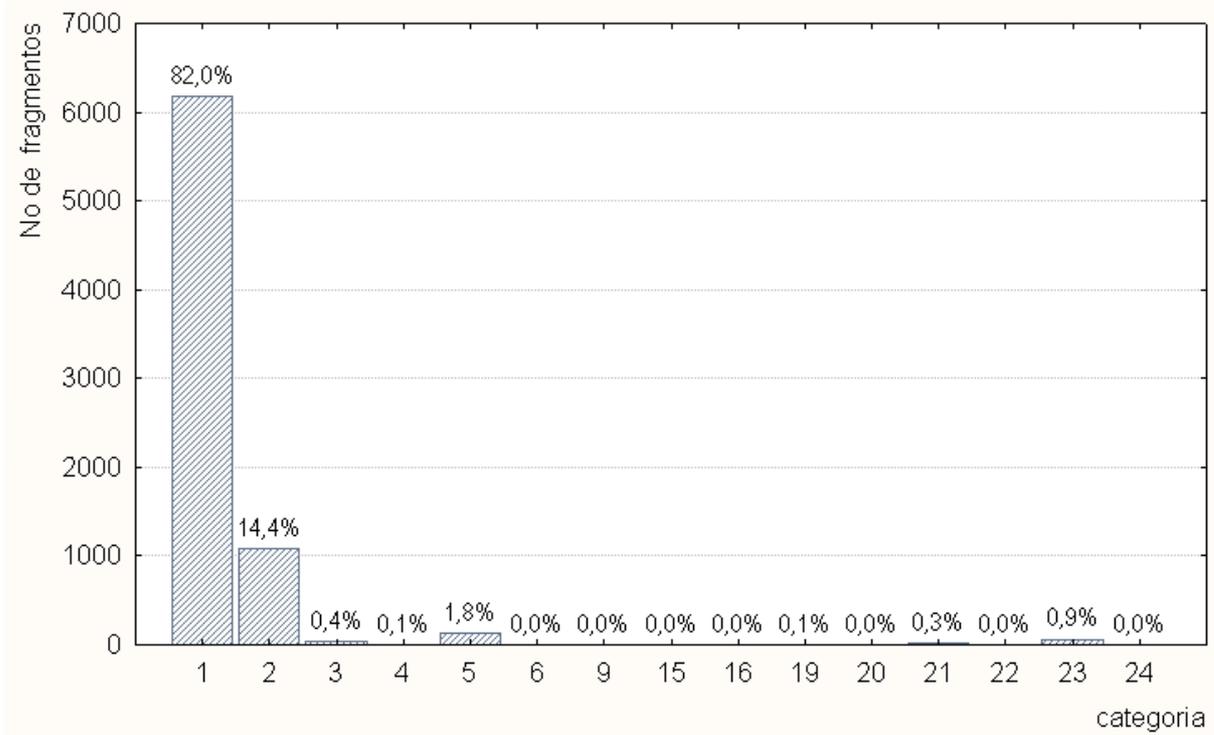


FIGURA 7 - Atributo quanto à origem (parte do vaso) e forma do fragmento – classe.

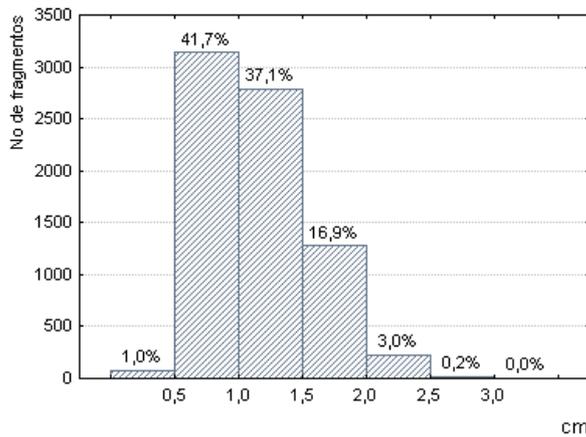
As categorias 6 (parede e base) - fragmentos 4392 e 4405, 9 (asa) – fragmento 270, 15 (rolete de confecção) - fragmento 10295, 16 (ombro) - fragmento 3974, 20 (borda com furo de suspensão) - fragmento 8182, 22 (polidor de sulco) – fragmentos 9283, 9287 e 10677 e 24 (sem descrição) representam menos de 0,1% do total de fragmentos registrados.

Palavra da especialista:

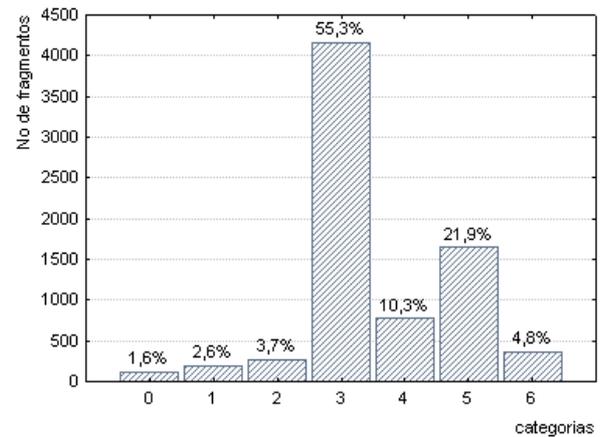
A presença de ombro (ou parede angular) é um forte indicador da Tradição Tupiguarani. A presença de roletes de confecção ou bolotas de argila indica áreas de confecção de vasos. Os chamados furos de suspensão têm sido encontrados em pequena quantidade na área do Vale do Rio Paranapanema, Estados de São Paulo e Paraná, e servem para suspender a peça ou para reparar rachaduras. A quantidade de polidores de sulco em cerâmica, na área do ProjPar varia de um Sítio para o outro. Fragmentos que apresentam base e parede são raros, pois no geral as peças do Sítio

Piracanjuba estão bastante fragmentadas. De modo geral, podemos afirmar que todas as categorias de atributos, ora comentadas, são raras e aparecem no Sítio Piracanjuba e em outros sítios da área do ProjPar em pequena quantidade se compararmos com o número de fragmentos de parede, que constitui a maior parte do vaso.

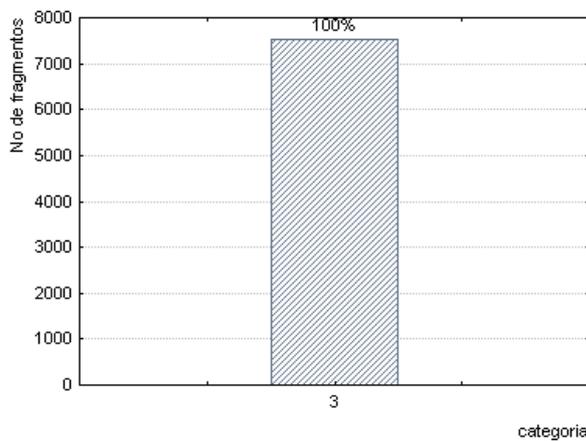
Quanto à fabricação ilustrada por seus atributos na figura 8: (a) espessura da parede, (b) queima, (c) dureza, (d) tratamento superficial, e na figura 9: (a) decoração interna, (b) decoração externa, (c) técnica de manufatura, pode-se dizer que:



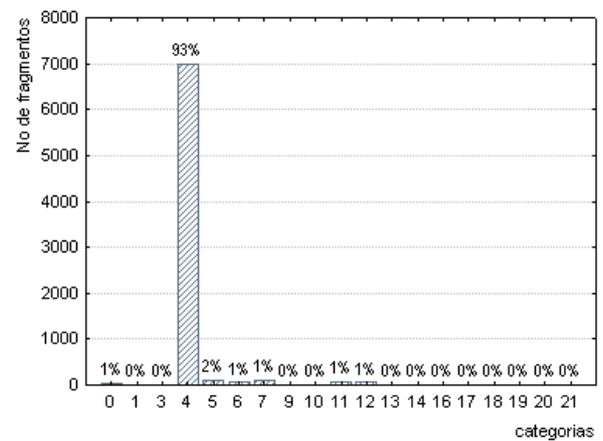
(a) Espessura da parede



(b) Queima



(c) Dureza



(d) Tratamento superficial

FIGURA 8 - Atributos quanto à fabricação: (a) espessura da parede, (b) queima, (c) dureza, (d) tratamento de superfície.

- a espessura da parede de 41,7% dos fragmentos varia entre 0,5 e 1,0 cm, 37,1% entre 1,0 e 1,5 cm, 16,9% entre 1,5 a 2,0 cm, sendo que fragmentos com espessura acima de 2,0 cm e abaixo de 0,5 cm perfazem 3,2% e 1,0% do total, respectivamente;

- a queima de 55,3% dos fragmentos apresenta seção transversal com presença do núcleo central escuro e uma camada interna e externa clara (categoria 3), 21,9% seção transversal com uma camada clara na parte externa e uma camada escura na interna (categoria 5), 10,3% seção transversal sem presença de núcleos, com cor uniforme variando do cinza escuro ao preto (categoria 4), 4,8% seção transversal com uma camada clara na parte interna e uma camada escura na externa (categoria 6); 3,7% seção transversal sem presença de núcleos, com cor uniforme variando do cinza claro ao pardo (categoria 2); 2,6% seção transversal sem presença de núcleos, com cor uniforme variando do laranja tijolo ao amarelo (categoria 1) e 1,5% das peças não estão identificadas quanto à queima (categoria 0);

- todos os fragmentos se enquadram no nível 3 da escala de dureza de Mohs, a escala-padrão de dureza estabelecida por Mohs tem por parâmetro dez minerais: talco, gipso, calcita, fluorita, apatita, ortoclásio, tóquio, cáridon e diamante.

Palavra da especialista: “No atributo dureza (DUREZ) foi constatada a não variação. Uma explicação possível para este tipo de ocorrência seria o uso de uma mesma fonte de argila”.

Portanto, este atributo será desincorporado do estudo;

- o tratamento superficial de 93% das peças é alisado interna e externamente (categoria 4), 2% polido internamente e alisado externamente (categoria 5); 1% alisado internamente e polido externamente (categoria 6), 1% polido interna e externamente (categoria 7), 1% pertencem à categoria 11 sem descrição na categorização, 1% à categoria 12 sem descrição na categorização e 1% sem identificação (categoria 0). As demais categorias aparecem em pequenas proporções não perfazendo isoladamente 1% dos fragmentos coletados.

Quanto à decoração interna, ilustrada na figura 9 (a), 73% dos fragmentos possuem decoração lisa (categoria 1), 12% engobo branco (categoria 11), 6% engobo laranja (categoria

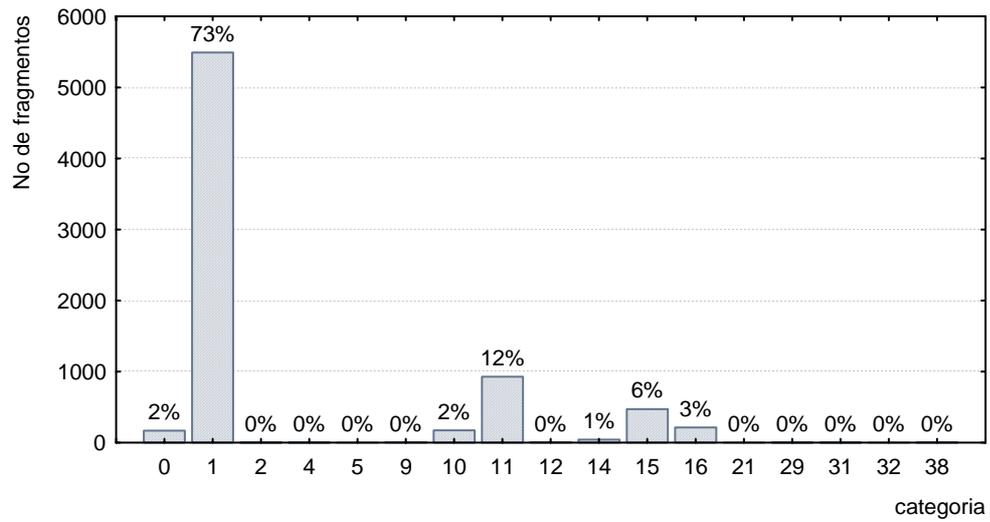
15), 3% pintado (categoria 16), 2% engobo vermelho (categoria 10), 2,0% sem identificação (categoria 0) e 1% engobo vermelho/branco (categoria 14).

As demais categorias presentes: inciso (categoria 4), corrugado (categoria 5), engobo preto (categoria 9), engobo preto/vermelho (categoria 12), engobo branco/inciso no contorno da garganta (categoria 21), pintado/engobo branco (categoria 29), engobo vermelho pintado (categoria 32) e sem descrição (categoria 38), não alcançam a marca de 1% dos fragmentos individualmente.

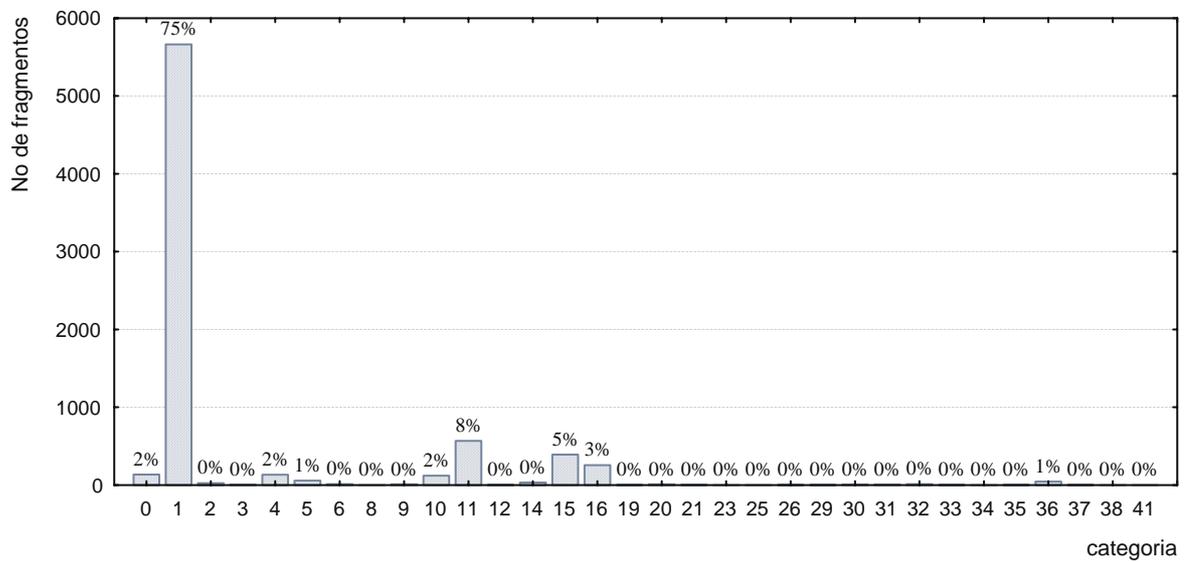
A decoração externa dos fragmentos coletados, ilustrada na figura 9 (b), apresenta grande diversidade, 32 tipos diferenciados, sendo que em 75% dos mesmos aparece a decoração externa lisa (categoria 1), em 8% engobo branco (categoria 11), em 5% engobo laranja (categoria 15), em 3% pintado (categoria 16), em 2% sem identificação (categoria 0), em 2% inciso (categoria 4), em 2% engobo vermelho (categoria 10) e em 1% sem descrição (categoria 36). As demais categorias não perfazem individualmente 1% dos fragmentos coletados.

Quanto à técnica de manufatura, ilustrada na figura 9 (c), 99,5% dos fragmentos coletados foram confeccionados com a técnica do roletado (acordelado) (categoria 1), 0,5% modelado a mão (categoria 3). Tendo sido registrado 1 (um) fragmento moldado (categoria 4).

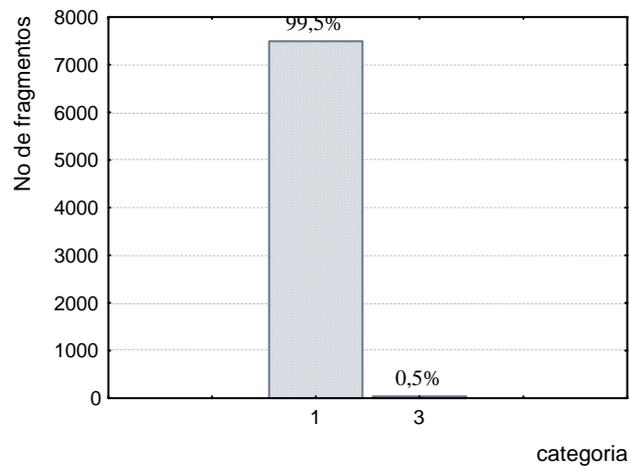
Quanto ao estado de conservação, ilustrado na figura 10, 96,4% dos fragmentos encontra-se em bom estado de conservação (categoria 2), 2,9% pouco conservados (categoria 3) e 0,7% bastante conservados (categoria 1).



(a) Decoração Interna



(b) Decoração Externa



(c) Técnica de Manufatura

FIGURA 9 - Atributos quanto à fabricação: (a) decoração interna, (b) decoração externa, (c) técnica de manufatura.

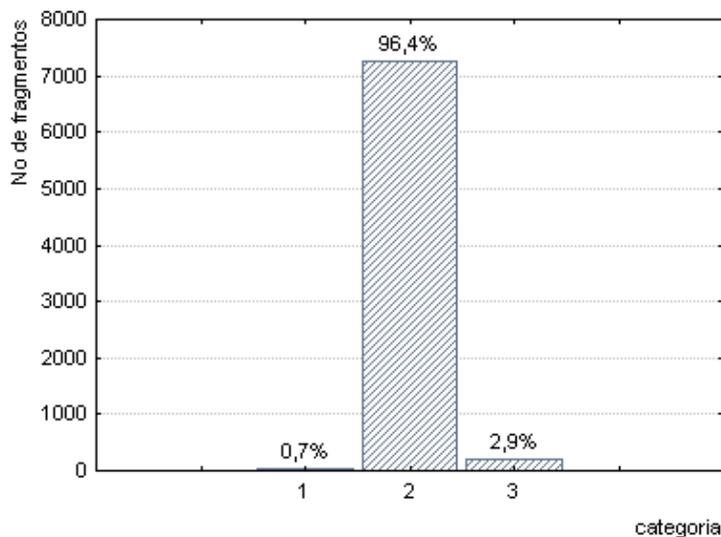


FIGURA 10 - Estado de conservação

1.10.5 Conclusões da análise preliminar dos dados e atributos selecionados.

A análise preliminar dos dados indica que os fragmentos em sua maioria, 96,4%, encontram-se em bom estado de conservação, tendo como material constituinte preponderante o material mineral com granulometria entre 0,1 e 0,26mm + caco moído com granulometria menor que 0,1 mm, sendo que a espessura da parede em quase 42% dos casos varia entre 0,5 e 1,0 cm.

As queimas mais comuns, encontradas em 77% dos fragmentos, são as que apresentam seção transversal com presença do núcleo central escuro e uma camada interna e externa clara (preponderante) e a que apresenta seção transversal com uma camada clara na parte externa e uma camada escura na interna.

A decoração lisa é a mais encontrada tanto interna quanto externamente, sendo que dentre das demais categorias de decoração a de engobo branco é a mais presente, seguida da de engobo laranja. O tratamento de superfície preponderante é o alisado interna e externamente, e a técnica de manufatura em 99,5% das peças é a do roletado.

Quanto aos atributos pode-se dizer que sua distribuição não se ajusta à curva normal, sendo em sua grande maioria categorizados.

Os atributos selecionados constam no quadro 8 e a descrição das categorias adotada de Moraes e Faccio (2004), no quadro 9.

Quadro 8 - Atributos selecionados para o processo da Descoberta de Conhecimento.

Especificação	Atributo	Abrev.	Tipo	Codificação
Origem (parte do vaso)	Classe	CL	Qualitativo categórico	1 – 6, 9, 15, 16, 19 – 24
Material Constituinte	Antiplástico	TP	Qualitativo categórico	1, 13 e 19
	Material Mineral	MN	Quantitativo contínuo	–
	Caco Moído	CM	Quantitativo contínuo	–
	Carvão	CV	Quantitativo contínuo	–
Fabricação	Espessura da Parede	EP	Quantitativo contínuo	–
	Queima	Q	Qualitativo categórico	0 – 6
	Tratamento Superficial	TS	Qualitativo categórico	0, 1, 3 – 7, 9 – 14, 16 – 21
	Decoração Interna	DI	Qualitativo categórico	0, 1, 2, 4, 5, 9 – 12, 14 – 16, 21, 23, 31, 32, 38
	Decoração Externa	DE	Qualitativo categórico	0 – 6, 8 – 12, 14 – 16, 19 – 21, 23, 25, 26, 29, 30 – 38, 41
	Técnica de Manufatura	MAN	Qualitativo categórico	1, 3, 4

Fonte: MORAIS e FACCIO (2004).

Quadro 9 - Descrição dos atributos_categorias selecionados para análise.

Atributo	Abrev.	Categoria	Codificação
Classe	CL	1	parede
		2	borda
		3	base
		4	base, parede, borda
		5	parede angular
		6	parede e base
		9	asa
		15	rolete de confecção
		16	ombro
19	parede com furo de suspensão		

continua

continuação

Atributo	Abrev.	Categoria	Codificação
Classe	CL	20	borda com furo de suspensão
		21	borda c/ suporte para tampa
		22	polidor de sulco
		23	borda/parede angular
		24	sem descrição
Antiplástico	TP	1	mineral
		13	mineral e caco moído
		19	mineral, carvão e caco moído
Material Mineral	MN	-	-
Caco Moído	CM	-	-
Espessura da Parede	EP	-	-
Queima	Q	0	não identificada
		1	seção transversal sem presença de núcleos, com cor uniforme, variando do laranja tijolo ao amarelo
		2	seção transversal sem presença de núcleos, com cor uniforme variando do cinza claro ao pardo
		3	seção transversal com presença de núcleo central escuro e uma camada interna e externa clara
		4	seção transversal sem presença de núcleos, com cor uniforme variando do cinza escuro ao preto
		5	seção transversal com uma camada clara na parte externa e uma camada escura na interna
		6	seção transversal com uma camada clara na parte interna e uma camada escura na externa
Tratamento Superficial	TS	0	não identificado
		1	sem tratamento externo/ interno
		3	alisamento interno sem alisamento externo
		4	alisamento interno e externo
		5	polimento interno/ alisamento externo
		6	polimento externo/ alisamento interno
		7	polimento interno e externo
		9	lustro externo/ alisamento interno
		10	lustro interno/ alisamento externo
		Tratamento Superficial	TS
13	brunidura/ polimento externo		
14	polimento interno/ sem tratamento externo		
16	polimento externo/ sem tratamento interno		
17, 18, 19, 20 e 21	sem descrição		
Decoração Interna e Decoração Externa	DI e DE	0	não identificada
		1	liso
		2	entalhado
		3	ungulado
		4	inciso
		5	corrugado
		6	escovado
		8	pinçado
		9	engobo preto
		10	engobo vermelho
		11	engobo branco
		12	engobo preto/vermelho
		14	engobo vermelho/branco
		15	engobo laranja
		16	pintado

continua

conclusão

Atributo	Abrev.	Categoria	Codificação
Decoração Interna e Decoração Externa	DI e DE	19	digitado
		20	serrungulado
		21	engobo branco/inciso no contorno da garganta
		23	corrugado associado ao ungulado
		25	engobo branco associado ao inciso
		26	pintado associado ao inciso
		29	pintado/ engobo branco
		30	roletado
		31	engobo branco/laranja
		32	engobo vermelho/pintado
	33, 34, 35, 36, 37 e 41	sem descrição	
Técnica de Manufatura	MAN	1	roletado (acordelado)
		3	modelado à mão
		4	moldado

Fonte: MORAIS e FACCIIO (2004).

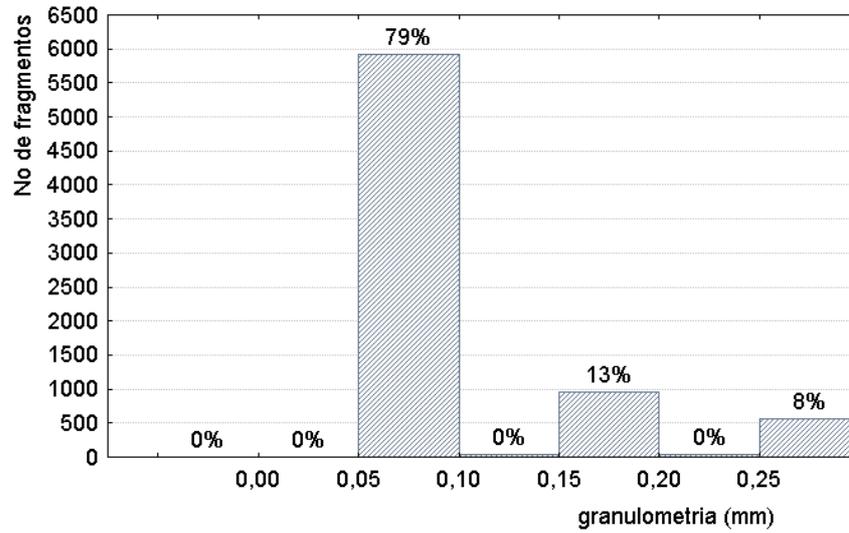
1.10.6 Tratamento dos dados

Dentre os atributos analisados nesta fase, material mineral, caco moído, carvão e espessura da parede não são categorizados. Considerando a necessidade da adequação desses atributos à técnica de mineração de dados eleita e análise de correspondência, efetivou-se a categorização das mesmas.

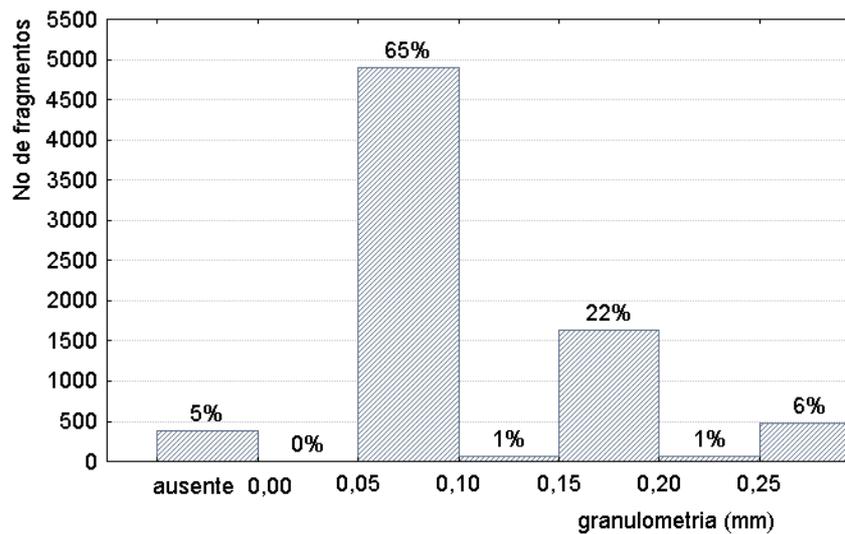
A análise dos histogramas do material mineral e do caco moído ilustrados na figura 11 (a) e (b), respectivamente, revela que existe para esses atributos uma categorização implícita, talvez imposta involuntariamente pelas limitações dos equipamentos utilizados para a medição ou por limitação dos sentidos humanos. A categorização é ilustrada no quadro 10, no qual o código 0 para o caco moído significa ausência, ou seja, antiplástico da categoria 1.

O atributo espessura da parede foi categorizado pelos quartis 1 (Q1) e 3 (Q3) e pela mediana (Md). Assim, para valores iguais ou menores que Q1 foi atribuída a categoria 1, maior que Q1 e menor ou igual à Md, categoria 2, maior que Md e menor ou igual à Q3, categoria 3, e maior que Q3, categoria 4. Devido à baixa frequência do atributo carvão a categorização não considerou a granulometria do material, apenas a sua presença, código 1 e

ausência, código 0. A categorização desses atributos encontra-se ilustrada no quadro 10.



(a) granulometria do material mineral



(b) granulometria do caco moído

FIGURA 11 - Histogramas da granulometria dos materiais constituintes: material mineral (a), caco moído (b).

Quadro 10 - Categorização dos atributos material mineral, caco moído, carvão e espessura da parede.

material mineral		caco moído		carvão		esp. da parede	
(mm)	cat.	(mm)	cat.	(mm)	cat.	(cm)	cat.
0,0 — 0,1	1	ausente	0	ausente	0	≤ 0,9	1
0,1 — 0,2	2	0,0 — 0,1	1	presente	1	0,9 — 1,1	2
> 0,2	3	0,1 — 0,2	2			1,1 — 1,5	3
		> 0,2	3			> 1,5	4

1.11 MINERAÇÃO DE DADOS (DATA MINING)

Mineração de dados é o passo que transforma dados em informações e caracteriza-se pela definição e aplicação da técnica e do algoritmo que diante de uma tarefa específica seja capaz de extrair eficientemente conhecimento explícito e útil de um conjunto de dados (AURÉLIO, 1999).

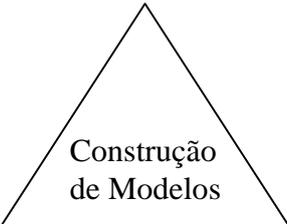
1.11.1 Tarefas de mineração de dados

Entende-se aqui tarefa como o conjunto de problemas que têm características similares através de diferentes domínios de aplicação (MOXON, 1996). Por exemplo, as mesmas tarefas utilizadas para desenvolver a mineração de dados na detecção de fraudes de um banco, podem ser usadas para detectar “bons pagadores” no comércio. Em geral, a diferença reside então no domínio da aplicação, no repositório de dados, nos atributos e como eles são utilizados na análise.

As tarefas de mineração de dados são especificadas quanto ao nível de informações que o perito deseja sobre o domínio de aplicação, incluindo desde a visualização dos dados até a construção de modelos preditivos (KOPERSKI et al., 1996), conforme ilustrado na figura 12.

Portanto, para cada tarefa de mineração de dados que diferem nos tipos de problemas que pretendem resolver, existe um conjunto de técnicas capazes de extrair os relacionamentos relevantes dos dados:

O quadro 11 apresenta as principais tarefas realizadas por técnicas de mineração de dados.



Construção
de Modelos

Fonte: Adaptado de KOPERSKI et al. (1996).

FIGURA 12 - Níveis de informação do domínio da aplicação.

Quadro 11 - Tarefas de mineração de dados e as técnicas associadas a elas.

TAREFAS	TÉCNICAS
Regras de associação	Estatística e Teoria dos Conjuntos
Classificação	Alg. Genéticos, Redes Neurais e Árvores de Decisão
Agrupamento	Redes Neurais e Estatística
Previsão de séries temporais	Redes Neurais, Lógica Nebulosa e Estatística

Fonte: AURÉLIO (1999).

Regras de associação: consiste em determinar o conjunto de atributos que tendem a coocorrer. O exemplo clássico é determinar quais produtos costumam ser colocados juntos em um carrinho de compras de supermercado, daí o termo análise de “*market basket*”. As cadeias de varejo usam associação para planejar a disposição dos produtos nas prateleiras das lojas ou em um catálogo, de modo que os itens geralmente adquiridos na mesma compra sejam vistos próximos entre si.

“A tarefa de associação pode ser considerada uma tarefa bem definida, determinística e relativamente simples, que não envolve predição da mesma forma que a tarefa de classificação” (FREITAS, 2000).

Classificação: A tarefa de classificação consiste em construir um modelo que possa ser aplicado a um conjunto de dados visando categorizá-los em classes, ou seja, um objeto é examinado e classificado de acordo com uma classe definida (HARRISON, 1998). Groebel e

Gruenwald (1999) citam como exemplo de tarefa de classificação: classificar pedidos de créditos em baixo, médio e alto risco.

Agrupamento (ou Clustering): é um processo de partição de uma população heterogênea em subgrupos ou “*clusters*” mais homogêneos (HARRISON, 1998). Exemplos de agrupamento: agrupar os clientes por região do país, agrupar clientes com comportamento de compra similar (GROEBEL e GRUENWALD, 1999).

É importante fazer aqui uma distinção entre agrupamento e classificação de dados.

Um agrupamento é um processo pelo qual uma amostra da população é segmentada de acordo com algum critério de similaridade.

Uma classificação envolve a atribuição de rótulos às amostras de uma população, de acordo com algum critério externo, em geral baseado nas convenções e preceitos humanos.

A figura 13 ilustra a diferença entre agrupamento e classificação.

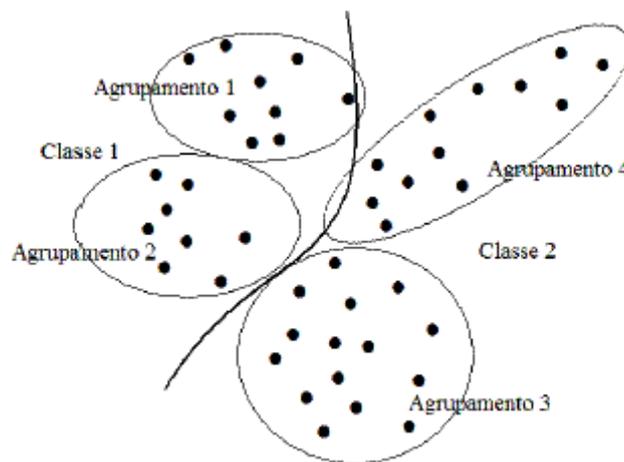


FIGURA 13 - Agrupamento X Classificação

Um estudo interessante a ser feito sempre que se usa um método de agrupamento (não supervisionado) é verificar se ele fornece uma divisão dos dados em grupos que tenham alguma similaridade com as classes em que nós, humanos, os classificamos.

Em princípio, pode não haver qualquer relação entre os agrupamentos e as classes, e nesse caso o agrupamento pode estar revelando novas associações entre os dados, que podem

ou não ser úteis, possibilitando a AQUISIÇÃO DE CONHECIMENTO. Porém, se for possível interpretar os agrupamentos em termos de classes já conhecidas, então o método de agrupamento pode estar fazendo a tarefa de classificação.

De um ponto de vista teórico, o agrupamento é uma forma não-paramétrica de estimação de densidade de pontos.

Previsão de séries temporais: a análise de “*market basket*” tradicional trata de uma coocorrência de atributos. Uma variante deste problema ocorre quando há uma informação adicional para amarrar junto com uma seqüência de ocorrência dos atributos (como por exemplo, em uma compra de supermercado, um número de cliente, um cartão de crédito) em uma série de tempo. Nesta situação, além da coocorrência dos atributos também a ordem e o intervalo de tempo é importante.

As regras que capturam estes relacionamentos podem ser usadas, no exemplo, para identificar um jogo típico das compras do precursor que podem predizer a compra subsequente de um artigo específico. No cuidado de saúde, tais métodos podem ser usados para identificar custos de tratamento rotineiros e excepcionais, bem como o tempo excedente dos procedimentos múltiplos.

A escolha das técnicas de mineração de dados é função da tarefa específica a ser executada e dos dados disponíveis para a análise, sendo que a familiaridade com as técnicas é indispensável para a melhor abordagem do problema apresentado (FAQ, 1996).

1.11.2 Técnicas estatísticas

As técnicas estatísticas estão dentre as mais utilizadas na mineração de dados, pois é uma área do conhecimento já sedimentada e amplamente estudada, disponibilizando aos

usuários grande número de algoritmos já implementados e otimizados (KOPERSKI; HAN e ADHIKARY, 1998).

Segundo Von Wangenheim (2004), do ponto de vista da estatística, as técnicas úteis para reconhecimento e descoberta de padrões em ambientes nos quais os fenômenos são descritos/baseados em uma grande variedade de dados são conhecidas como Análise Exploratória de Dados (AED) ou estatística exploratória, dentre elas encontram-se as técnicas de agrupamento e as de redução de dimensionalidade.

1.11.3 Agrupamento e redução de dimensionalidade

O objetivo principal das técnicas de agrupamento ou aglomeração é separar objetos ou observações em agrupamentos naturais de forma que os elementos pertencentes a um mesmo grupo tenham um alto grau de semelhança ou similaridade, enquanto que, quaisquer elementos pertencentes a grupos distintos tenham pouca ou nenhuma similaridade (ANDEBERG, 1973).

A similaridade entre dois objetos ou observações pode ser expressa como uma função da distância entre dois pontos representativos dos mesmos no espaço n-dimensional. A maneira mais usual de calcular a distância entre dois pontos i e j no espaço n-dimensional é conhecida por Distância Euclidiana (d_{ij}) e é dada por (01):

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (01)$$

sendo:

d_{ij} a distância (similaridade) entre os objetos i e j;

x_{ik} o valor do atributo k para o objeto i;

x_{jk} o valor do atributo k para o objeto j .

Quando todos os atributos são numéricos os agrupamentos são facilmente detectados pela distância Euclidiana ou por Mahalanobis. Segundo Neves do Vale (2006), a distância de Mahalanobis considera que as superfícies de cada classe são elipsóides centrados na média, sendo que para o caso especial em que a variância é a mesma para todos os atributos e a covariância é zero as superfícies são esferas e a distância de Mahalanobis equivale à distância Euclidiana. Porém, quando os atributos são expressos por dados binários ou qualitativos o agrupamento por meio de medidas de distância é bastante complicado e nem sempre conduz a bons resultados (HAN, 2001).

Muitas vezes, um pequeno número destes atributos contém as informações mais relevantes, enquanto que o restante dos atributos adiciona pouco ou nada à interpretação dos resultados.

A redução de atributos por meio de critérios objetivos, permitindo a sua visualização em gráficos contendo maior informação estatística, pode ser conseguida por meio da análise fatorial, análise de componentes principais e da análise de correspondência. A análise fatorial e a análise de componentes principais exigem dados mensurados em escala numérica contínua (ordinal) não sendo apropriadas para dados nominais (LANDIM, 2001).

Nesses casos, a análise de correspondência, uma técnica exploratória que analisa a distribuição de massa de um conjunto de observações, configura-se como a mais adequada.

A análise de correspondência apresenta as relações geométricas do cruzamento, contingenciamento, de atributos qualitativos categóricos em um conjunto de pontos capaz de ser representado graficamente por um pequeno número de dimensões (redução de dimensionalidade) (PEREIRA, 2004), de forma que a proximidade entre os pontos representados está relacionada com o nível de associação entre esses atributos (agrupamento) (FIGUERAS e GARGALLO, 2003).

1.11.4 Análise de correspondência ou análise das associações.

O termo análise de correspondência vem do francês “Analyse Factorielle de Correspondances” utilizado por Benzécri e outros que desenvolveram a técnica (BARTHOLOMEW et al., 2002).

Segundo Chaib Filho (2004), o objetivo da análise de correspondência é o estudo das:

- relações existentes no interior de um conjunto X, ou seja, o estudo das relações entre os atributos de X;
- relações existentes no interior de um conjunto Y, ou seja, o estudo das relações entre os atributos de Y;
- relações existentes entre os elementos do conjunto X e os elementos do conjunto Y, ou seja, o estudo das relações existentes entre os atributos de X e Y.

O ponto de partida para a análise de correspondência é uma tabela de contingência como a ilustrada na figura 14.

Essas tabelas são constituídas de dados relativos ao número de elementos existentes nos atributos combinados de X e Y,

X/Y	y ₁	...	y _j	...	y _c	FREQÜÊNCIA MARGINAL DE X
X ₁	n ₁₁	n _{1c}	n _{1.}
...
X _i	n _{i1}	...	n _{ij}	...	n _{ic}	n _{i.}
...
X _r	n _{r1}	...	n _{rj}	...	n _{rc}	n _{r.}
FREQÜÊNCIA MARGINAL DE Y	n _{.1}	...	n _{.j}	...	n _{.c}	n _{..}

FIGURA 14 - Tabela de contingência.

sendo:

X – conjunto de atributos $\{x_1, \dots, x_r\}$;

Y – conjunto de atributos $\{y_1, \dots, y_c\}$;

$n_{i.}$ – total de unidades de x_i em Y;

$n_{.j}$ – total de unidades de y_j em X;

$$\text{FREQUÊNCIA MARGINAL DE X} - n_{i.} = \sum_j n_{ij} ;$$

$$\text{FREQUÊNCIA MARGINAL DE Y} - n_{.j} = \sum_i n_{ij} ;$$

$$\text{FREQUÊNCIA TOTAL} - n_{..} = \sum_i \sum_j n_{ij}$$

Na análise fatorial de correspondência não são considerados os valores absolutos, mas as correspondências entre as frequências relativas isto é, os valores relativos cuja soma é igual a 1, ou seja, os valores originais são transformados de modo a poder ser interpretados como probabilidades condicionais.

Dessa forma uma primeira caracterização das distribuições de linhas e colunas pode ser obtida pelos:

1. perfis que descrevem as distribuições condicionadas, sendo:

- a. perfis linha descrevem as distribuições condicionadas de uma linha segundo as colunas, como em

$$\frac{n_{i1}}{n_{i.}}, \frac{n_{i2}}{n_{i.}}, \dots, \frac{n_{ij}}{n_{i.}}, \dots, \frac{n_{ic}}{n_{i.}} \quad (02)$$

- b. perfis coluna descrevem as distribuições condicionadas de uma coluna segundo as linhas

$$\frac{n_{1j}}{n_{.j}}, \frac{n_{2j}}{n_{.j}}, \dots, \frac{n_{ij}}{n_{.j}}, \dots, \frac{n_{rj}}{n_{.j}} \quad (03)$$

2. perfis marginais (massa) que descrevem a distribuição marginal:

a. das linhas condicionadas pelo total geral da tabela

$$\frac{n_{1.}}{n_{..}}, \frac{n_{2.}}{n_{..}}, \dots, \frac{n_{i.}}{n_{..}}, \dots, \frac{n_{r.}}{n_{..}} \quad (04)$$

b. das colunas condicionadas pelo total geral da tabela

$$\frac{n_{.1}}{n_{..}}, \frac{n_{.2}}{n_{..}}, \dots, \frac{n_{.j}}{n_{..}}, \dots, \frac{n_{.c}}{n_{..}} \quad (05)$$

3. perfis médios:

a. das linhas

$$\frac{n_{i.}}{r} \quad (06)$$

b. das colunas

$$\frac{n_{.j}}{c} \quad (07)$$

A validação da utilização da técnica da análise de correspondência em uma tabela de contingência é feita por meio do teste do chi-quadrado de Pearson, das comparações dos perfis linha e coluna com os perfis marginais correspondentes, expressão (08), determinando a probabilidade da associação global entre as linhas e colunas.

Para o teste: a hipótese nula H_0 é: X e Y são independentes, e a hipótese alternativa H_1 é: X e Y são dependentes. Convém observar que se H_0 é certa todos os perfis linha/coluna são iguais entre si e iguais ao perfil marginal das linhas/colunas,

$$G^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^r \sum_{j=1}^c \frac{n_{i.} \left(\frac{n_{ij}}{n_{i.}} - \frac{n_{.j}}{n_{..}} \right)^2}{\frac{n_{.j}}{n_{..}}} = \sum_{i=1}^r \sum_{j=1}^c \frac{n_{.j} \left(\frac{n_{ij}}{n_{.j}} - \frac{n_{i.}}{n_{..}} \right)^2}{\frac{n_{i.}}{n_{..}}} \quad (08)$$

sendo:

$$e_{ij} = E[n_{ij} | H_0 \text{ é certa}] = \frac{n_{i.} n_{.j}}{n_{..}}.$$

Valores pequenos de G^2 significam que os valores de n_{ij} e e_{ij} estão próximos e, portanto, que H_0 é certa. Pelo contrário, valores grandes evidenciam que H_0 é falsa. O valor p do teste é dado por

$$p\left[\chi^2_{(r-1)(c-1)} \geq G^2\right] \quad (09)$$

Para um nível de significância $0 < \alpha < 1$, rejeita H_0 se o valor de p for menor ou igual a α .

Se a hipótese nula for rejeitada, então a natureza das associações pode ser explorada examinando-se as porcentagens linhas e/ou colunas das tabelas de contingência. (BARTHOLOMEW et al., 2002).

O teste do chi-quadrado é hábil em determinar a probabilidade da associação global entre as linhas e colunas, mas não fornece nenhuma informação a respeito de quais são as associações individuais significativas entre pares de linhas/colunas da matriz de dados. Por outro lado, a análise de correspondência mostra como as linhas/colunas estão associadas/relacionadas, ou seja, o exame das razões específicas dos desvios da hipótese de independência é a razão de ser da análise de correspondência (FIGUERAS e GARGALLO, 2003). A técnica da análise de correspondência consiste essencialmente em encontrar a decomposição em valores singulares da matriz \mathbf{C} de dimensões $r \times c$, cujos elementos são:

$$\frac{n_{ij} - e_{ij}}{\sqrt{e_{ij}}} \quad (10)$$

em que:

n_{ij} – é o número de unidades na casela (i,j) da tabela de contingência, e

e_{ij} – representa o número esperado de unidades na casela (i,j) , calculado sob a hipótese de que

X e Y são independentes.

A partir da matriz \mathbf{C} , constrói-se um sistema de coordenadas associado às linhas e às colunas da tabela de contingência, que reflitam as relações existentes entre as linhas e as

colunas.

Neste sistema de coordenadas, as distâncias entre os perfis são calculadas pelas “distâncias χ^2 ”, que são distâncias pitagóricas ponderadas entre os perfis e dadas pelas expressões seguintes:

distância entre perfis linha

$$d_{ij}^2 = \sum_{k=1}^c \frac{1}{n_{.k}} \left(\frac{n_{ik}}{n_{.i}} - \frac{n_{jk}}{n_{.j}} \right)^2 \quad (11)$$

distância entre perfis colunas

$$d_{ij}^2 = \sum_{k=1}^r \frac{1}{n_{k.}} \left(\frac{n_{ki}}{n_{.i}} - \frac{n_{kj}}{n_{.j}} \right)^2 \quad (12)$$

A divisão de cada termo ao quadrado pela frequência esperada é a variação padronizada que compensam as grandes variações devidas às altas frequências e as pequenas variações em baixas frequências. Caso nenhuma padronização fosse realizada, as diferenças entre grandes proporções tenderiam a ser grandes e dominar o cálculo das distâncias, enquanto as diferenças entre pequenas proporções tenderiam a desaparecer. A ponderação dos fatores é utilizada para equalizar essas diferenças (NAGPAUL, 2006).

As distâncias χ^2 possuem a propriedade da equivalência distribucional, a qual afirma que se duas linhas/colunas de N são proporcionais, então o seu amalgamento não afetará as distâncias entre colunas/linhas, ou seja, as distâncias χ^2 não são sensíveis a variações na codificação das categorias com comportamento similar quanto a seus perfis condicionais. Por causa da natureza dessa transformação as relações entre colunas e linhas da tabela transformada são as mesmas que aquelas da matriz original de dados.

Portanto, a análise de correspondência busca encontrar 2 matrizes de coordenadas cartesianas tais que:

$$\mathbf{A} = \begin{bmatrix} a'_1 \\ \dots \\ a'_r \end{bmatrix} \text{ represente os pontos linha com } a'_i = (a_{i1}, \dots, a_{ic}) \text{ e}$$

$$\mathbf{B} = \begin{bmatrix} b'_1 \\ \dots \\ b'_c \end{bmatrix} \text{ represente os pontos coluna com } b'_j = (b_{1j}, \dots, b_{rj}).$$

Quando a dimensão da matriz **A** ou **B** for igual a três pode-se recorrer a um sistema triangular plano para a sua representação (PEREIRA, 2004), mas para dimensões maiores que três a análise de correspondência terá que recorrer a técnicas de redução de dimensionalidade com o intuito de derivar um sistema de coordenadas para projeções planas, de forma que o produto final mostre num espaço bidimensional definido pelos dois mais importantes autovetores a melhor representação dos dados, cujas posições dos pontos reflitam as associações/relações detectadas pelo teste do χ^2 .

Existem diversas maneiras de calcular (normalizar) as matrizes **A** e **B**, que dependem do tipo de análise que se quer realizar. Segundo Pereira (2004), a normalização é feita por uma distribuição da inércia pelas linhas, pelas colunas ou por ambas. Quando é feita pelas linhas ou colunas não se analisa concomitantemente pontos linha e pontos coluna, apenas as relações intrapontos linha/coluna. Quando é feita por ambas, normalização simétrica ou canônica, pode-se analisar as relações intrapontos linha/coluna, bem como as relações entre pontos linhas e colunas.

Pontos linha próximos indicam linhas com perfis similares e pontos coluna próximos indicam colunas com perfis similares (WICHERN e JOHNSON, 2001).

Cabe ressaltar que não se pode interpretar as distâncias entre pontos linha e pontos coluna, mas é legítimo interpretar a posição relativa de um ponto linha/coluna com todos os outros pontos linha/coluna.

A normalização canônica decompõe a matriz \mathbf{C} em valores singulares, calculando as matrizes \mathbf{U} , \mathbf{D} e \mathbf{V} , tais que:

$$\mathbf{C} = \mathbf{UDV}' \quad (13)$$

sendo:

$$\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}$$

\mathbf{U} – é uma matriz de dimensão $r \times k$

\mathbf{V} - é uma matriz de dimensão $c \times k$

$$k = \min \{r-1, c-1\}$$

\mathbf{D} – é a matriz diagonal contendo os valores singulares $(\mu_1, \mu_2, \dots, \mu_k)$, ordenados do maior para o menor.

As matrizes \mathbf{A} e \mathbf{B} são calculadas a partir de (14) e (15)

$$\mathbf{A} = \mathbf{D}_r^{-1/2} \mathbf{U} \mathbf{D} \quad (14)$$

$$\mathbf{B} = \mathbf{D}_c^{-1/2} \mathbf{V} \mathbf{D} \quad (15)$$

nas quais:

$$\mathbf{D}_r = \text{diag}(n_1, \dots, n_r) \text{ e } \mathbf{D}_c = \text{diag}(n_1, \dots, n_c).$$

1.11.4.1 Inércia

A inércia aqui é o “momento de inércia” da mecânica. Um objeto físico tem um centro de gravidade (centróide) e cada partícula do objeto tem uma massa m e uma determinada distância d do centróide. O momento de inércia do objeto é o somatório da massa multiplicada pela distância ao quadrado de todas as partículas integrantes do objeto.

$$\sum md^2 \quad (16)$$

Este conceito encontra analogia na análise de correspondência, para tanto considera-se que há uma nuvem de pontos dos perfis, sendo:

1. a nuvem de pontos dos perfis das linhas: o conjunto de pontos cujas coordenadas são os componentes dos perfis de linha e cujas massas são os componentes do perfil marginal das linhas;
2. a nuvem de pontos dos perfis das colunas: o conjunto de pontos cujas coordenadas são os componentes dos perfis de coluna e cujas massas são os componentes do perfil marginal das colunas.

Esses pontos têm um centróide (perfil médio) e uma distância (χ^2) entre pontos do perfil. Cada ponto do perfil contribui à inércia da nuvem inteira e pode ser computada por

1. para o ponto do perfil de linha da *i*-ésima linha

$$\frac{1}{n_{..}} \sum_j \frac{n_i \left(\frac{n_{ij}}{n_i} - \frac{n_{.j}}{n_{..}} \right)^2}{\frac{n_{.j}}{n_{..}}} \quad (17)$$

2. para o ponto do perfil de coluna da *j*-ésima coluna

$$\frac{1}{n_{..}} \sum_i \frac{n_j \left(\frac{n_{ij}}{n_j} - \frac{n_{i.}}{n_{..}} \right)^2}{\frac{n_{i.}}{n_{..}}} \quad (18)$$

A inércia total da tabela de contingência é dada por

$$\frac{1}{n_{..}} \sum_{i=1}^r \sum_{j=1}^c \frac{n_i \left(\frac{n_{ij}}{n_i} - \frac{n_{.j}}{n_{..}} \right)^2}{\frac{n_{.j}}{n_{..}}} = \frac{1}{n_{..}} \sum_{i=1}^r \sum_{j=1}^c \frac{n_j \left(\frac{n_{ij}}{n_j} - \frac{n_{i.}}{n_{..}} \right)^2}{\frac{n_{i.}}{n_{..}}} = \frac{G^2}{n_{..}} \quad (19)$$

e quantifica quanto de variação existe nos perfis de linha ou de coluna. É a partir dela que se calculam as proporções de inércia explicada a cada dimensão (20) bem como as proporções

de inércia acumulada explicada pelas i primeiras dimensões (21).

$$IT = \sum_{k=1}^K \mu_k^2 = \sum_{k=1}^K \sum_{i=1}^r n_{i.} a_{ik}^2 = \sum_{k=1}^K \sum_{j=1}^c n_{.j} b_{jk}^2$$

$$\left\{ \frac{\mu_i^2}{IT}; i = 1, \dots, K \right\} \quad (20)$$

$$\left\{ \sum_{k=1}^i \frac{\mu_k^2}{IT}; i = 1, \dots, K \right\} \quad (21)$$

Contribuições Totais – medem a importância dos pontos linha/coluna na construção dos eixos fatoriais. Para interpretar os eixos são utilizados os atributos com contribuições mais significativas.

Contribuição i -ésima linha

$$Ct_k(i) = \frac{n_{i.} a_{ik}^2}{\sum_{j=1}^r n_{.j} a_{jk}^2} = \frac{n_{i.} a_{ik}^2}{\mu_k^2} \quad (22)$$

Contribuição j -ésima coluna

$$Ct_k(j) = \frac{n_{.j} b_{jk}^2}{\sum_{i=1}^c n_{.i} b_{ik}^2} = \frac{n_{.j} b_{jk}^2}{\mu_k^2} \quad (23)$$

sendo:

$$\sum_{i=1}^r Ct_k(i) = \sum_{j=1}^c Ct_k(j) = 1$$

Contribuições relativas – medem a qualidade de representação de um ponto a cada dimensão e é medida pela distância desse ponto, o quadrado do co-seno do ângulo entre a dimensão k -ésima e o ponto linha i -ésima/coluna j -ésima, à origem das coordenadas do sistema derivado. Esse valor pode ser interpretado como a correlação do ponto com as dimensões derivadas (GREENACRE, 1984).

Contribuição relativa i -ésima linha

$$Cr_k(i) = \frac{a_{ik}^2}{\sum_{i=1}^k a_{i1}^2} \quad (24)$$

Contribuição relativa j -ésima coluna

$$Cr_k(j) = \frac{b_{kj}^2}{\sum_{j=1}^k b_{1j}^2} \quad (25)$$

sendo:

$$\sum_{i=1}^r Cr_k(i) = \sum_{j=1}^c Cr_k(j) = 1$$

Qualidade da representação de um ponto linha/coluna em um determinado número de dimensões derivadas representa a contribuição desse ponto à inércia total explicada por esse determinado número de dimensões, e é dada pelo somatório das contribuições relativas desse ponto linha/coluna a essas dimensões.

Elementos Suplementares – são linhas ou colunas da tabela de contingência não utilizadas no cálculo dos eixos fatoriais, porém uma vez calculadas são plotadas no gráfico com a finalidade de ajudar na interpretação dos resultados obtidos. Suas coordenadas são calculadas utilizando-se as relações baricêntricas existentes entre os pontos linha e coluna.

Tabela de resíduos padronizados – Quando o teste do χ^2 detecta a existência de relações/associações em uma tabela de contingência é interessante a investigação da tabela de resíduos.

Os resíduos padronizados

$$P_{resij} = \frac{n_{ij} - e_{ij}}{\sqrt{e_{ij}} \sqrt{\left(1 - \frac{n_{i.}}{n_{..}}\right) \left(1 - \frac{n_{.j}}{n_{..}}\right)}} \quad (26)$$

sendo:

n_{ij} = frequência observada;

e_{ij} = frequência esperada;

$n_{.i}$ = total da coluna;

$n_{.j}$ = total da linha;

$n_{..}$ = total geral da tabela,

apresentam os resíduos (diferenças entre o valor observado e o esperado) em uma forma padronizada expressos em unidades de desvio padrão, podendo instruir pontos de corte para um nível de significância de excesso ou falta de ocorrências. (HABERMAN, 1978).

Segundo Pereira (2004) “os resíduos padronizados representam valores de relação biunívoca com probabilidades de ocorrência, isto é, valores maiores que 1,96 ou menores que -1,96 têm pequenas chances de ocorrência (+ ou - 2,5%)”, e podem instruir pontos de corte para a falta ou excesso de ocorrências, permitindo a distinção entre as casuais e as causais que são a causa das associações/relações detectadas pelo teste do χ^2 .

1.11.4.2 Sugestões úteis para a interpretação dos resultados da análise de correspondência

- pontos linha/coluna podem ter uma boa qualidade de representação em um determinado número de dimensões derivadas, mas baixa contribuição à inércia total, ou seja, um ponto linha/coluna com um perfil similar ao perfil marginal de linha/coluna;
- um ponto faz uma elevada contribuição à inércia de uma dimensão em duas situações, quando tem uma grande distância do baricentro, mesmo se tiver uma massa pequena, ou quando tem uma massa grande e uma distância pequena;

- pontos perto da origem têm uma distribuição proporcionalmente homogênea ao perfil marginal;
- pontos do conjunto de linhas situados longe da origem mas perto entre si tem perfis similares. O mesmo raciocínio se aplica ao conjunto dos pontos coluna.

É comum a apresentação dos pontos linha e pontos coluna em um único gráfico de dispersão, no qual pode-se interpretar a distância dos pontos linha entre si, a distância dos pontos coluna entre si e a distância entre um determinado ponto de um conjunto (uma linha ou coluna) com todos os outros pontos do outro conjunto.

Pereira (2004) propõe que a análise entre um ponto linha/coluna e o conjunto dos outros pontos coluna/linha se desenvolva considerando a dimensão de projeção que passa pelo ponto linha/coluna e a origem das dimensões derivadas: *“A partir do centro do sistema de coordenadas podem-se traçar retas que ligam a origem a qualquer um dos pontos representados. Essas retas são dimensões de projeção para o exame das relações entre objetos com base no ponto para o qual se deriva a reta.”*

Micheloud (1997) amplia a base de interpretação quando considera o ângulo formado pelas dimensões derivadas que passam pela origem e pelos pontos linha/coluna em estudo.

Para tanto, quando o ângulo é agudo ($<90^0$) então o ponto linha/coluna e o ponto coluna/linha estão correlacionados positivamente, se o ângulo é obtuso ($>90^0$) a correlação é negativa e se for reto ($=90^0$) a correlação não se estabelece. A figura 15 ilustra a regra.

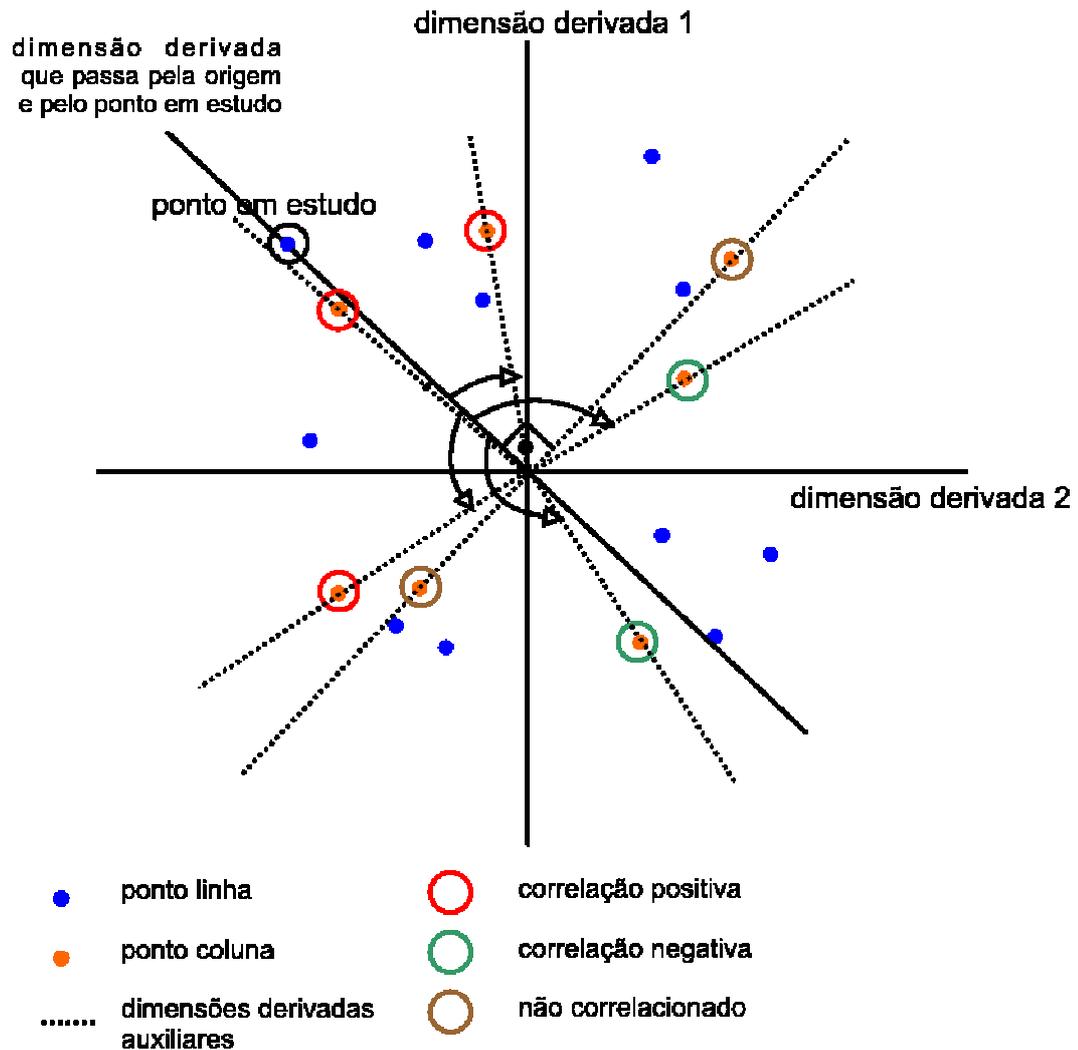


FIGURA 15 - Regras geométricas auxiliares para a interpretação do gráfico de dispersão dos resultados da análise de correspondência

1.12 A QUESTÃO ESPACIAL

Considerando que estacas e núcleos antropogênicos são a representação simbólica de um atributo espacial, muito deste trabalho se perderia se esse fator espacial não fosse levado em consideração.

Desta maneira, para premiar a questão espacial optou-se pela geração de modelos numéricos de terreno (MNT) dos resíduos padronizados dos atributos_categorias que apresentam falta ou excesso de ocorrências.

O MNT apresentado sob a forma de mapas temáticos gerados pelo fatiamento da

grade para um ponto de corte em um nível de significância de 2,5% e 5% dos resíduos padronizados, fornece um instrumento de análise visual privilegiado do comportamento espacial dos atributos_categorias dos fragmentos cerâmicos.

Cabe aqui clarificar alguns conceitos e técnicas concernentes à modelagem numérica de terreno.

1.12.1 Conceitos básicos de Modelo Numérico de Terreno (MNT)

Segundo Felgueiras (1998): “Um Modelo Numérico de Terreno (MNT) é uma representação matemática computacional da distribuição de um fenômeno espacial que ocorre dentro de uma região da superfície terrestre, onde a superfície em geral é contínua e o fenômeno é variado”.

O objetivo de um MNT é a geração de uma superfície, por interpolação ou aproximação, representativa das características espaciais do fenômeno tendo como ponto de partida sua amostragem.

Segundo Silva (2004), o objetivo dos processos de interpolação e dos de aproximação é o mesmo, encontrar uma categoria de funções que seja a expressão mais aproximada de uma função capaz de descrever a dependência funcional de suas variáveis.

A diferença entre os dois processos reside no fato de que para os processos de interpolação busca-se a função $P(x)$ que passe pelos pontos $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$, sendo $y_0 = f(x_0), y_1 = f(x_1), \dots, y_n = f(x_n)$; enquanto que para os processos de aproximação busca-se a função aproximada $P(x)$ que se ajuste aos pontos $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$, segundo um método de aproximação escolhido (Campos, 1994).

Os processos de ajuste de superfície podem ser globais ou locais.

“Os métodos locais utilizam apenas um conjunto de pontos mais próximos ao retalho para o qual se deseja ajustar a superfície. Os métodos globais ajustam uma única superfície utilizando todos os pontos amostrados.” (NAMIKAWA et al.,2003).

Os modelos de ajuste locais utilizam funções de interpolação locais definidas por critérios de distância e/ou vizinhança das localizações geográficas e dos atributos na construção do reticulado (grade regular ou triangular) para aproximação da superfície.

Segundo Landim (2002):

“...os métodos mais comuns de interpolação são: Vizinho mais Próximo, Vizinho Natural, Triangulação Linear, Triangulação de Delaunay, Polígonos de Voronoi, Inverso da Potência da Distância (sendo o IQD, Inverso do Quadrado da Distância o mais utilizado), Mínima Curvatura, Regressão polinomial, Krigagem (com mais de 30 técnicas distintas) e Máxima Entropia Bayesiana.”

Segundo Felgueiras (1998), a média das cotas das amostras vizinhas (média móvel) se apresenta como uma das técnicas de interpolação mais simples para estimar os valores de cota dos pontos de uma grade regular retangular, sua formulação geral se encontra em

$$Z_i = \frac{\sum_{j=1}^n w_{ij} z_{ij}}{\sum_{j=1}^n w_{ij}} \quad (27)$$

sendo:

z_i o valor de cota de um ponto i qualquer da grade;

z_{ij} a cota de uma amostra j vizinha do ponto i da grade;

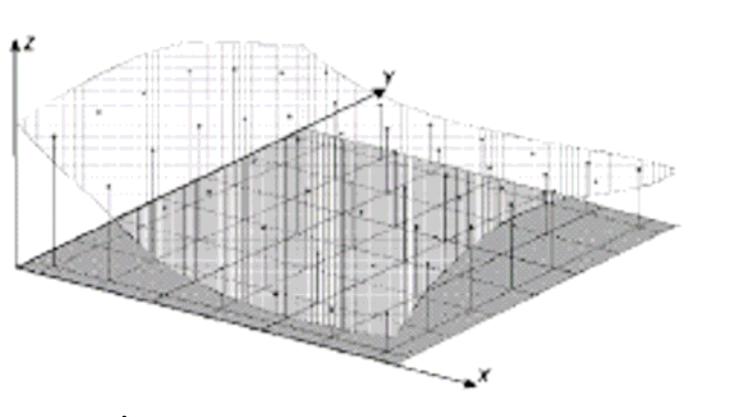
w_{ij} um fator de ponderação;

n número total de amostras consideradas.

As variações do fator de ponderação w_{ij} conduzem a interpoladores bastante conhecidos e já citados. Por exemplo, quando w_{ij} é dado por $1/n$ – média aritmética, por $1/d$ sendo d a distância entre as amostras vizinhas e o ponto a ser interpolado – inverso da distância, por $1/d^2$ – inverso do quadrado da distância.

A grade regular aproxima a superfície por meio de um poliedro de faces retangulares

(FELGUEIRAS e CÂMARA, 2001), ilustrada na figura 16. Os vértices desses poliedros não são necessariamente os pontos amostrados, salvo quando as amostras já foram adquiridas coincidentemente com os vértices da grade.

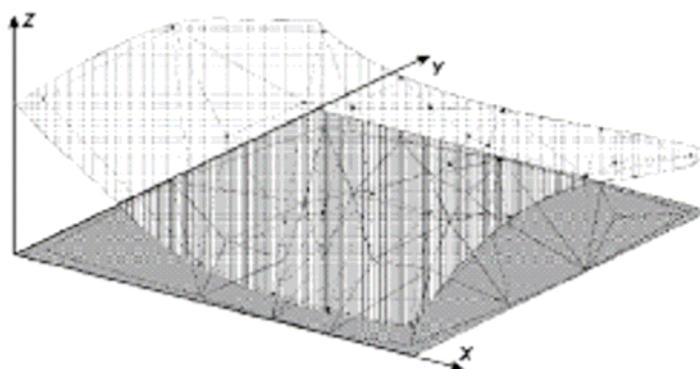


Fonte: FELGUEIRAS, C. A. e CÂMARA, G. (2001).

FIGURA 16 - Grade retangular regular

A grade triangular ilustrada na figura 17 é formada pelo conjunto de poliedros cujas faces são triângulos e os vértices do triângulo os pontos amostrados da superfície (FELGUEIRAS e CÂMARA, 2001). Segundo Namikawa et al. (2003), dentre as maneiras de gerar os triângulos da grade triangular, a mais usada é a regra conhecida como triangulação de Delaunay.

Landim (2000) aponta como uma das vantagens dos métodos de interpolação pela média móvel, mais especificamente o inverso da distância em potências baixas, na geração da grade regular o destaque dado às anomalias locais. Considerando que esse destaque é desejado na visualização do excesso ou falta de fragmentos cerâmicos do atributo_categoria detectado pelo χ^2 , optou-se pela construção de mapas temáticos obtidos pelo fatiamento da grade regular retangular gerada pelo interpolador do inverso da distância para a apresentação da espacialização dos resíduos padronizados.



Fonte: FELGUEIRAS, C. A. e CÂMARA, G. (2001).

FIGURA 17 - Grade triangular

Segundo Felgueiras (1998), fatiamento é a definição de intervalos, ou fatias de cotas da grade na qual cada intervalo de cotas é associado a um tema ou classe com a finalidade de gerar uma imagem temática a partir de um modelo (grade). Os intervalos de fatiamento deste trabalho têm como ponto de corte a falta ou excesso de ocorrências de fragmentos dos atributos_categorias em nível de significância de 2,5% e 5% e constam do quadro 12.

Quadro 12 - Intervalos de fatiamento para a ocorrência de excesso/falta de atributos_categorias de fragmentos cerâmicos

Classe	Nível de significância	Cota de corte	Intervalo
Falta 2,5%	2,5%	- 1,96	< -1,96
Falta 5%	5%	- 1,64	$\geq -1,96$ e < -1,64
Esperado	-	-	$\geq -1,64$ e $\leq 1,64$
Excesso 5%	5%	1,64	> 1,64 e $\leq 1,96$
Excesso 2,5%	2,5%	1,96	> 1,96

Todas as grades regulares e os mapas temáticos apresentados neste trabalho são passíveis de consulta dinâmica com a ferramenta de informações do software SPRING. O banco de dados Piraju, o projeto denominado Piracanjuba e seus planos de informações encontram-se em mídia digital descrita no apêndice A.

1.13 ESTRATÉGIAS DE MINERAÇÃO.

Segundo Landim (2001), a técnica da análise das correspondências é bastante susceptível à presença de valores anômalos (“outliers”). Quando da presença dos mesmos os primeiros eixos apresentam esses valores acumulados em uma das extremidades do eixo e os demais agrupados na outra.

Como a análise preliminar dos dados indica categorias de atributos com frequências menores que 5%, a opção adotada para evitar a acumulação foi separar os fragmentos cerâmicos em tabelas a serem analisadas separadamente. O critério de corte estabelecido atinge registros que possuem ao menos uma categoria de atributo que contribui com menos de 10% da distribuição de frequência desse atributo.

Este critério de corte foi aplicado à tabela contendo todos os registros limpos e tratados e sucessivamente aos registros remanescentes do corte anterior propiciando a geração de 4 novas tabelas denominadas comum, raros, raríssimos e excepcionais obedecendo à ordem de corte, cujas especificações ilustram o quadro 13 e constam do apêndice A (mídia digital).

O procedimento adotado em contrapartida ao simples descarte dos registros com atributos cujas categorias apresentaram baixa frequência deve-se à importância que possuem em pesquisas arqueológicas.

Quadro 13 - Características das tabelas para análise.

Nome da tabela	Número de fragmentos	Locais de ocorrência	Atributos	Abreviatura do Atributo	Categorias
Comum	5319	115	Classe	CL	1 e 2
			Antiplástico	TP	1, 13
			Mat. Mineral	MNC	1 – 3
			Caco Moído	CMC	0 – 3
			Esp. da Parede	EPC	1 – 4
			Queima	Q	3 – 6

			Deco. Interna	DI	1, 11, 15
			Deco. Externa	DE	1, 11, 15
			Trat. Superficial	TS	4
			Téc. de Man.	MAN	1
Raros	1456	110	Classe	CL	1, 2, 5
			Antiplástico	TP	1, 13
			Mat. Mineral	MNC	1 – 3
			Caco Moído	CMC	0 – 3
			Esp. da Parede	EPC	1 – 4
			Queima	Q	0 – 6
			Deco. Interna	DI	0, 1, 10, 11, 15, 16
			Deco. Externa	DE	0, 1, 4, 10, 11, 15, 16
			Trat. Superficial	TS	4, 5, 7
			Téc. de Man.	MAN	1
Raríssimos	497	71	Classe	CL	1 – 3, 23
			Antiplástico	TP	1, 13
			Mat. Mineral	MNC	1 – 3
			Caco Moído	CMC	0 – 3
			Esp. da Parede	EPC	1 – 4
			Queima	Q	0 – 6
			Deco. Interna	DI	0, 1, 11, 15
			Deco. Externa	DE	0, 1, 2, 5, 10, 11, 14 – 16, 36
			Trat. Superficial	TS	0, 4, 6, 11, 12, 19, 20
			Téc. de Man.	MAN	1, 3
Excepcionais	292	52	Classe	CL	1 – 6, 9, 15, 16, 19 – 24
			Antiplástico	TP	1, 13, 19
			Mat. Mineral	MNC	1 – 3
			Caco Moído	CMC	0 – 3
			Esp. da Parede	EPC	1 – 4
			Queima	Q	0 – 6
			Deco. Interna	DI	0 – 2, 4, 5, 9 – 12, 14 – 16, 21, 29, 31, 32, 38
			Deco. Externa	DE	0 – 6, 8 – 12, 14 – 16, 19 – 21, 23, 25, 26, 29 – 38, 41
			Trat. Superficial	TS	0, 1, 3 – 7, 9 – 14, 16 – 18, 20, 21
			Téc. de Man.	MAN	1, 3

A análise preliminar das tabelas de contingência mostra claramente a baixa frequência de algumas categorias de atributo (>5) na tabela excepcionais, o que segundo Pereira (2004) desautoriza a aplicação do teste do χ^2 e, por conseguinte, a análise de correspondência. Desta maneira, a tabela excepcionais será analisada visualmente por gráficos de colunas.

As tabelas comum, raros e raríssimos serão objetos de estudos individuais, nos

quais serão analisados e avaliados:

1. a aplicabilidade da técnica pelo teste do χ^2 ;
2. os perfis das estacas e NSAs/atributos_categorias e os perfis marginais das estacas e NSAs/atributos_categorias;
3. as relações/associações entre as estacas e NSAs/atributos_categorias detectadas pela inspeção visual dos gráficos de dispersão em 2 ou 3 dimensões derivadas;
4. as relações/associações entre as estacas e NSAs e os atributos_categorias detectadas pelos métodos geométricos propostos por Pereira (2004) e Micheloud (1997) em gráficos de dispersão conjunto das estacas e NSAs e atributos_categorias em duas dimensões derivadas;
5. a contigüidade espacial dos pontos de concentração/falta dos atributos_categorias em mapas temáticos dos resíduos padronizados dos atributos_categorias.

1.13.1.1 Tabela de contingência fcomum

O estudo preliminar da tabela de contingência fcomum, mostra que todos os fragmentos cerâmicos analisados foram confeccionados segundo a mesma técnica de manufatura, roletado, tendo como tratamento superficial o alisamento interno e externo, encontrando-se em sua grande maioria (98%) em bom estado de conservação.

O gráfico de linhas dos perfis linha e do perfil marginal das linhas da tabela fcomum, que consta do apêndice A (mídia digital) no arquivo perfis – planilha perfis linha comum, apresenta picos de alta para determinados atributos_categorias, indicando a existência de atributos_categorias dominantes, ou seja, encontrados em um número maior de fragmentos cerâmicos, são eles: CL_1, MNC_1, CMC_1, Q_3, DI_1 e DE_1.

A dispersão dos perfis linha em relação ao perfil marginal das linhas varia de 10 a 2 pontos percentuais, indicando falta de homogeneidade na distribuição de fragmentos de determinados atributos_categorias nas estacas e NSAs.

O gráfico de linhas dos perfis coluna e do perfil marginal das colunas da tabela fcomum, que consta do apêndice A (mídia digital) no arquivo perfis – planilha perfis coluna comum, apresentou um pico de alta para o NSA1 dominante que em princípio inviabilizou a análise dos perfis coluna. O fator determinante do pico de alta é o número de fragmentos coletados neste local, mais de 50% dos fragmentos sob análise são oriundos do NSA1.

No sentido de obter uma visualização melhor destes perfis, o perfil do NSA1 foi substituído por um perfil simulado com valores correspondentes a 10% dos valores originais. Isto posto, a análise dos perfis dos atributos_categorias revela a disparidade na coleta de fragmentos, sendo que o NSA1 e o NSA2 configuram-se como os dois maiores pontos de coleta. A dispersão em torno do perfil marginal das colunas indica que os fragmentos coletados possuem atributos_categorias vinculados aos pontos de coleta.

A inércia total do sistema é igual a 0,08653, o que equivale a dizer que a distribuição dos atributos_categorias e das estacas e NSAs se aproxima do centro de massa da distribuição, indicando homogeneidade.

Ainda assim, o valor do teste do χ^2 é igual a 3221,9, sendo significativo a 2,5% de significância (o valor crítico da distribuição χ^2 com 2508 graus de liberdade é 2371, $p = 0,00$), isto é, existem associações/relações entre as linhas e colunas, entre os elementos das linhas e entre os elementos das colunas passíveis de serem estudados com a aplicação da técnica da análise de correspondência com normalização canônica.

A tabela 1 apresenta os valores singulares, os autovalores, a inércia explicada a cada dimensão e a inércia explicada acumulada. As três primeiras dimensões contribuem com 45,33% (em negrito) da inércia total. Considerando o número de dimensões, vinte e duas, a

serem derivadas e analisando o gráfico dos autovalores de cada dimensão derivada, figura 18, conclui-se que as dependências observadas podem ser estudadas pela sua representação nessas três dimensões.

Tabela 1 - Valores singulares, autovalores e inércia – tabela fcomum.

Dimensão	Valores Singulares	Autovalores	Inércia Explicada a cada dimensão	Inércia explicada acumulada
1	0,142395	0,020276	23,43166	23,4317
2	0,107851	0,011632	13,44191	36,8736
3	0,085536	0,007316	8,45494	45,3285
4	0,082681	0,006836	7,89987	53,2284
5	0,078340	0,006137	7,09223	60,3206
6	0,074515	0,005552	6,41648	66,7371
7	0,066991	0,004488	5,18612	71,9232
8	0,062656	0,003926	4,53667	76,4599
9	0,060412	0,003650	4,21759	80,6775
10	0,058577	0,003431	3,96524	84,6427
11	0,056535	0,003196	3,69352	88,3362
12	0,052268	0,002732	3,15710	91,4933
13	0,047419	0,002249	2,59843	94,0918
14	0,044923	0,002018	2,33211	96,4239
15	0,040988	0,001680	1,94147	98,3653
16	0,037610	0,001415	1,63465	100,0000
17	0,000000	0,000000	0,00000	100,0000
18	0,000000	0,000000	0,00000	100,0000
19	0,000000	0,000000	0,00000	100,0000
20	0,000000	0,000000	0,00000	100,0000
21	0,000000	0,000000	0,00000	100,0000
22	0,000000	0,000000	0,00000	100,0000

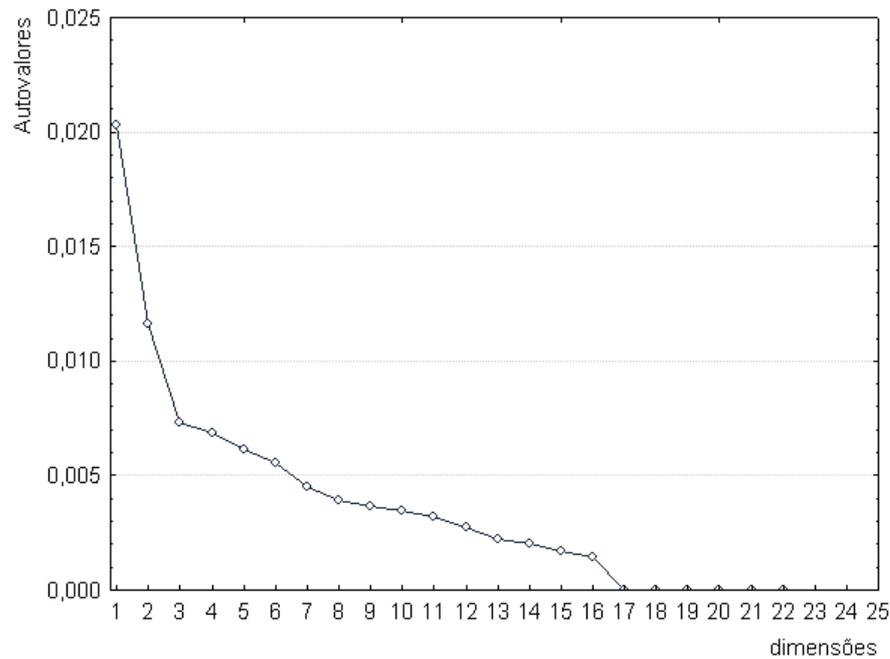


FIGURA 18 - Gráfico das dimensões derivadas – tabela fcomum

Encontra-se no apêndice A (mídia digital) o arquivo dimensões contendo as tabelas eletrônicas linhas comum, colunas comum, linhas raros, colunas raros, linhas raríssimos e colunas raríssimos que apresentam as coordenadas das três primeiras dimensões derivadas, os valores de massa, a qualidade da representação dos pontos, a contribuição à inércia total e a contribuição total e relativa para as tabelas fcomum, fraro e fraríssimos.

O exame da tabela linhas comum mostra que o NSA1 é quem mais contribui para a inércia da dimensão derivada 1 e tem a melhor qualidade de representação nesta dimensão derivada, o que equivale a dizer que o NSA1 se destaca dos demais NSAs bem como das estacas do sítio nesta dimensão. A tabela colunas comum revela que os atributos_categorias de granulometria do material constituinte, material mineral e caco moído, são os que mais contribuem para a inércia da dimensão 1 e apresentam a melhor qualidade de representação.

As dimensões derivadas 2 e 3 dão destaque a um conjunto de estacas que apresentam boa contribuição à inércia do sistema, apresentando boa qualidade de representação.

Os atributos_categorias que apresentam boa contribuição à inércia total e possuem boa qualidade de representação nas dimensões 2 e 3 são os de fabricação, mais

especificamente o de espessura de parede e os de decoração interna e externa

A análise dos gráficos de dispersão das dimensões derivadas 1, 2 e 3 das estacas e NSAs apresentadas duas a duas nas figuras 19 (dimensão 1X2), 20 (dimensão 1X3) e 21 (dimensão 2X3) da tabela fcomum, revela que:

- o NSA1 estabelece relação com a dimensão derivada 1;
- as estacas apresentam-se dispersas em uma forma elíptica com o raio maior paralelo à dimensão derivada 2, levemente deslocada em direção ao eixo positivo da dimensão derivada 1, indicando uma maior relação das estacas com a dimensão derivada 2;

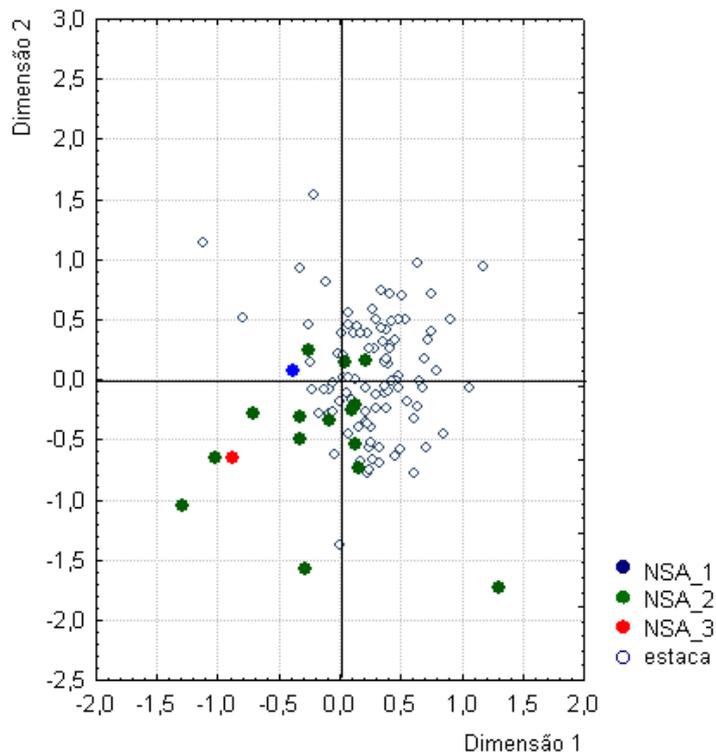


FIGURA 19 - Gráfico de dispersão das dimensões derivadas 1X2 das estacas e NSAs da tabela fcomum

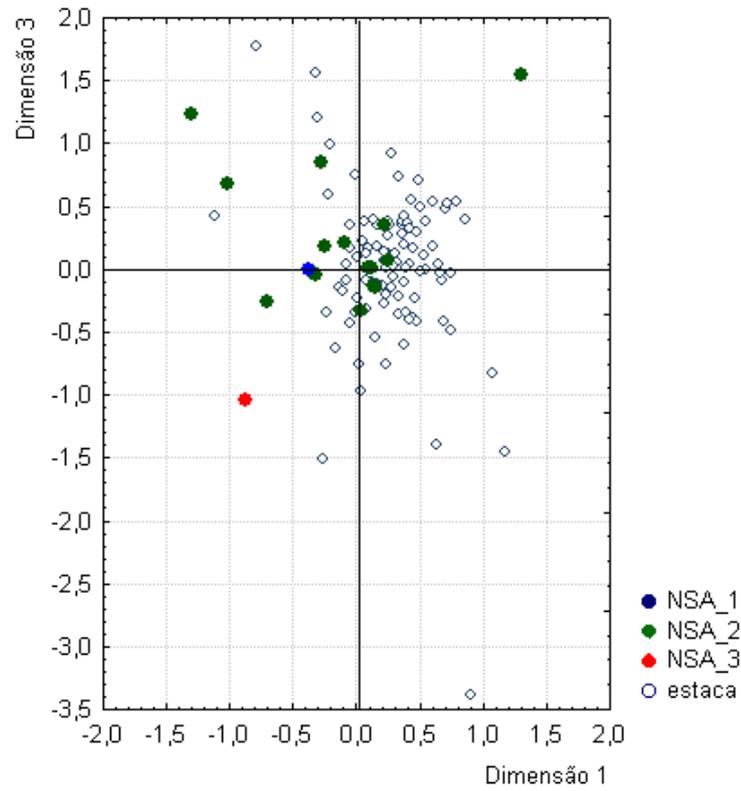


FIGURA 20 - Gráfico de dispersão das dimensões derivadas 1X3 das estacas e NSAs da tabela fcomum

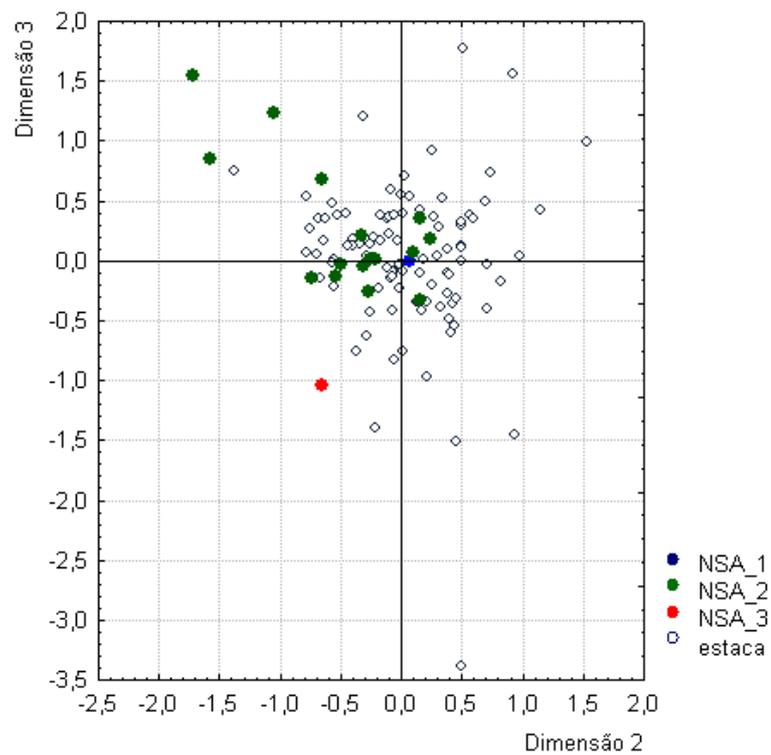


FIGURA 21 - Gráfico de dispersão das dimensões derivadas 2X3 das estacas e NSAs da tabela fcomum

- o NSA2 se apresenta bastante disperso em relação ao centro de massa da distribuição, onde se observa um conjunto de estacas do NSA2 apresentando um comportamento similar ao das estacas, ou seja, associado à dimensão derivada 2 e outro conjunto que se associa tanto à dimensão derivada 1 quanto à dimensão derivada 2;
- o NSA3 distancia-se do centro de massa, associando-se igualmente à dimensão derivada 1 e 2;

Quanto aos atributos_categorias apresentados nos gráficos de dispersão das dimensões derivadas 1, 2 e 3 duas a duas, 1X2, 1X3 e 2X3, nas figuras 22, 23 e 24, respectivamente, da tabela fcomum, pode-se dizer que:

- os atributos_categorias que estabelecem relações/associações com a dimensão 1 são: CMC_0-TP_1, CMC_2, CMC_3, MNC_2, MNC_3, MNC_1 e CMC_1, ficando claro que a dimensão 1 discrimina quanto ao material constituinte do fragmento cerâmico;

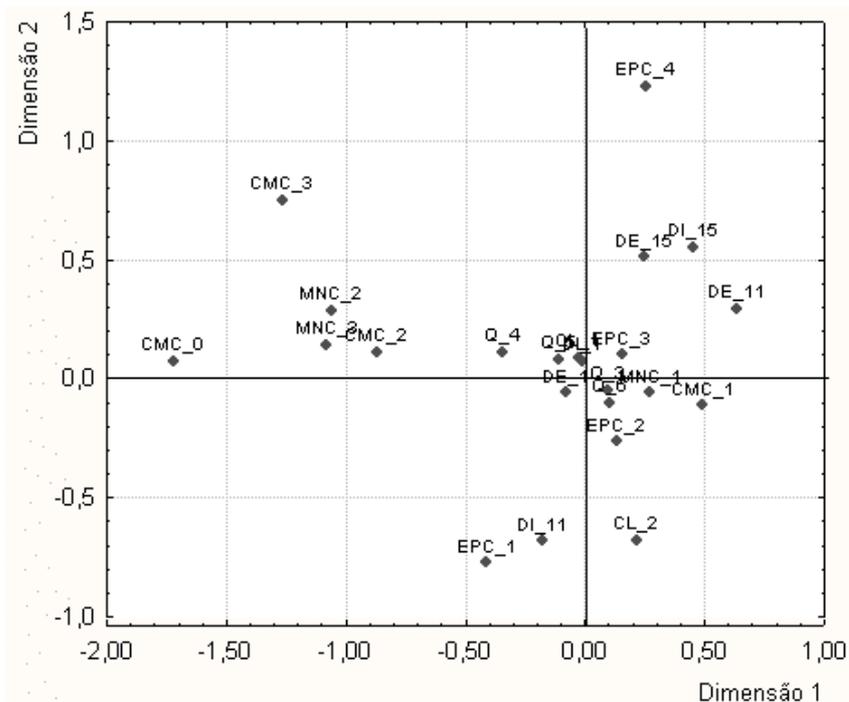


FIGURA 22 - Gráfico de dispersão das dimensões derivadas 1X2 dos atributos_categorias da tabela fcomum

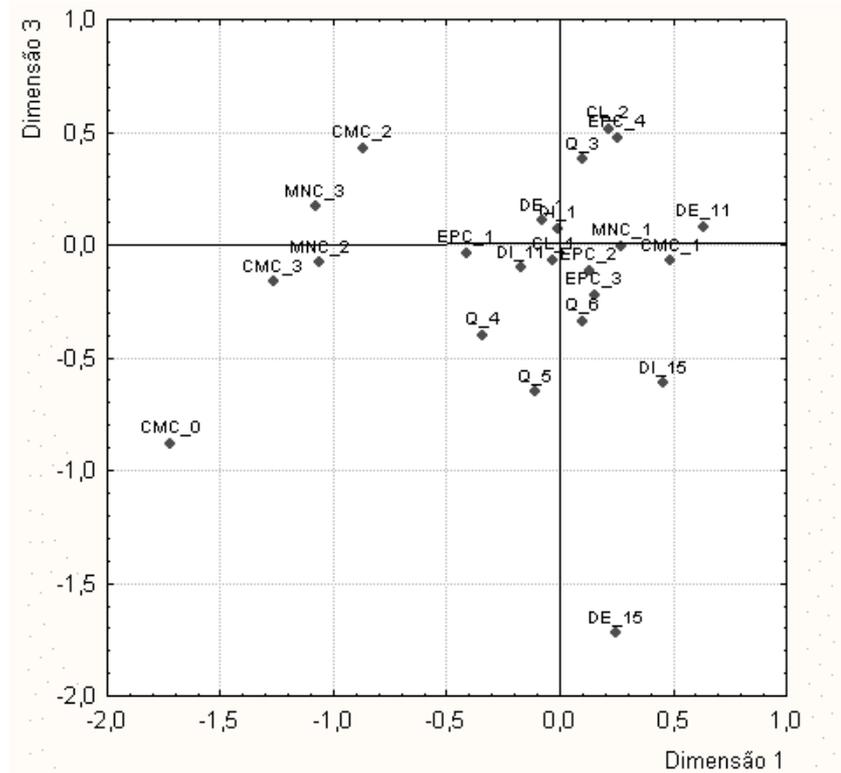


FIGURA 23 - Gráfico de dispersão das dimensões derivadas 1X3 dos atributos_categorias da tabela fcomum

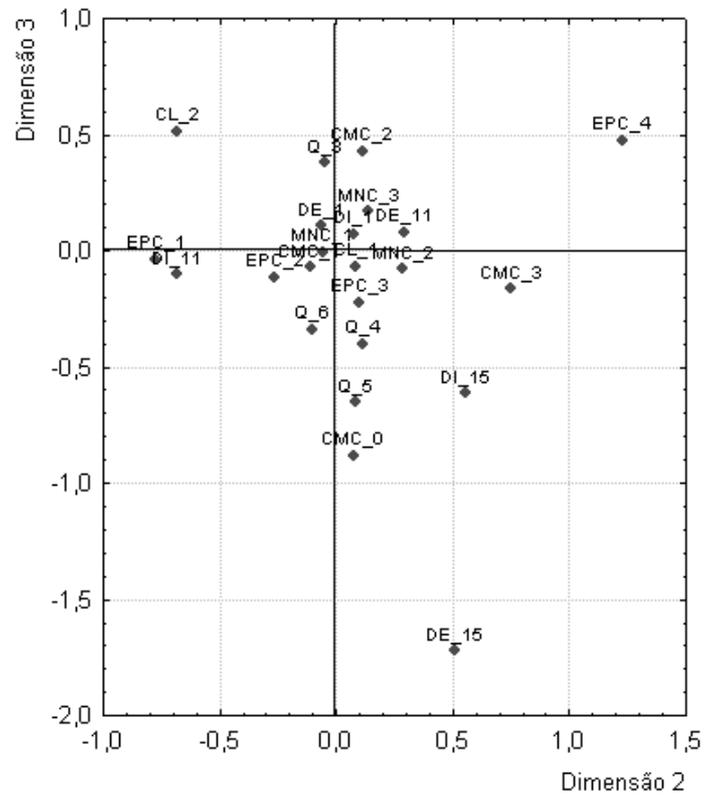


FIGURA 24 - Gráfico de dispersão das dimensões derivadas 2X3 dos atributos_categorias da tabela fcomum

- os atributos_categorias CL_2, EPC_1, EPC_4, DI_11, DI_15, DE_11 e DE_15 estabelecem relações/associações com a dimensão 2, podendo-se dizer que a dimensão 2 discrimina quanto à fabricação dos fragmentos cerâmicos;
- os atributos_categorias CL_1, DI_1, DE_1, EPC_2, EPC_3, Q_3, Q_5 e Q_6 se aproximam do centro de distribuição de massa, o que equivale a dizer que fragmentos cerâmicos que apresentam esses atributos_categorias foram coletados em todas as estacas e NSAs de forma proporcionalmente homogênea.

O exame da tabela de resíduos padronizados que consta do apêndice A (mídia digital) no arquivo de planilhas eletrônicas chamado resíduos_padronizados – planilha comum, revela não existir falta/excesso dos atributos_categorias CL_1, DI_1, DE_1 a um nível de significância de 5%. Para os demais, EPC_2, EPC_3, Q_3, Q_5 e Q_6, foram detectados pontos de falta/excesso em níveis de significância de 2,5%. A espacialização dos resíduos padronizados em modelos numéricos de terreno apresentados sob a forma de mapas temáticos para esses atributos_categorias é ilustrada nas figuras 25, 26, 27, 28 e 29, respectivamente, permitindo a visualização desses pontos de excesso/falta de ocorrência.

Os atributos_categorias CMC_0-TP_1, CMC_2, CM_3, MNC_2 e MNC_3 se afastam do centro de massa da distribuição indicando concentração em determinadas estacas e NSAs, em contrapartida ao MNC_1 e CMC_1 que se aproximam do centro de massa da distribuição, indicando certa homogeneidade proporcional na sua distribuição pelas estacas e NSAs.

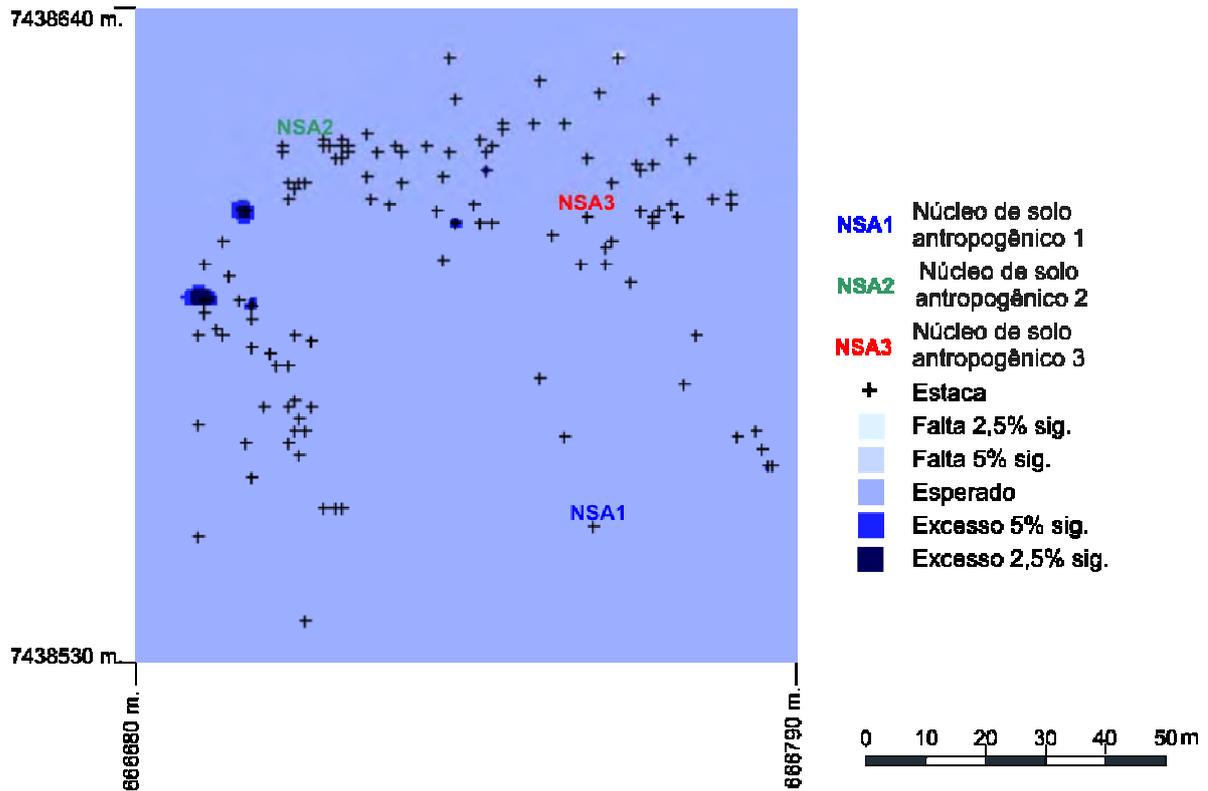


FIGURA 25 - Mapa temático do atributo espessura de parede categoria 2 (EPC_2).

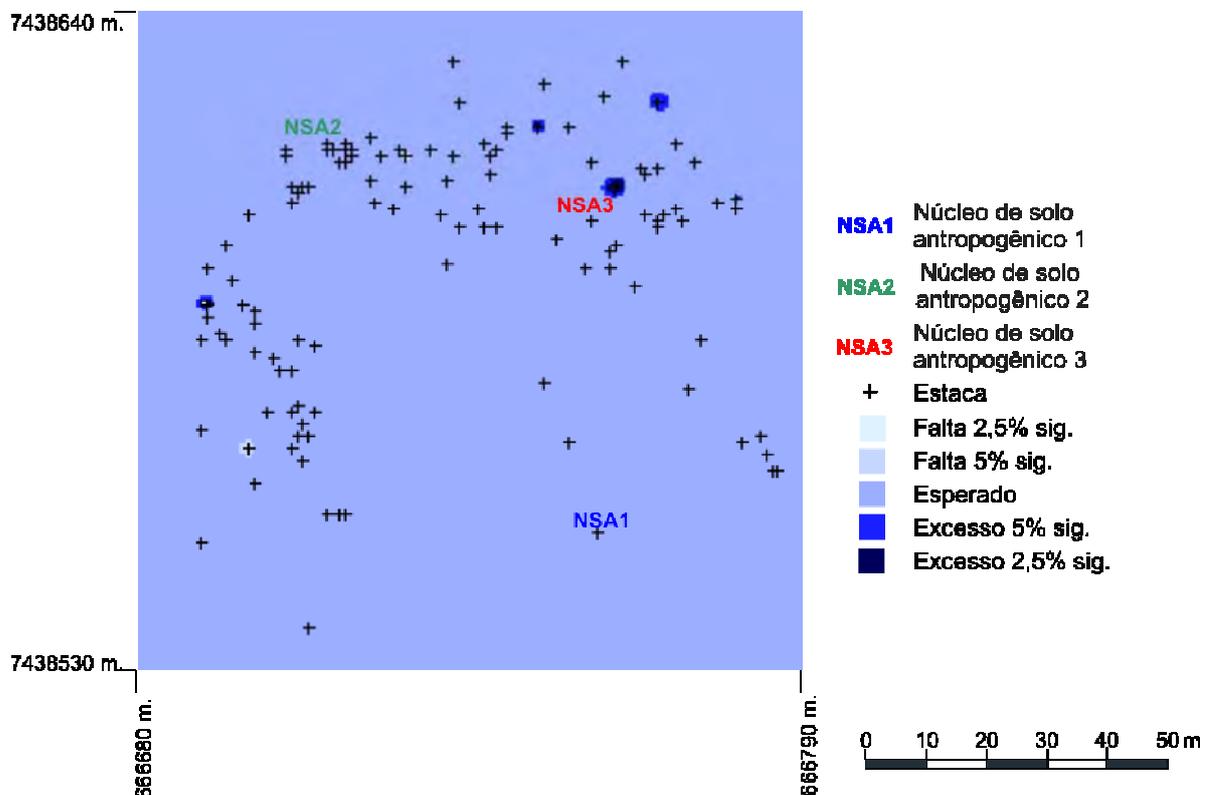


FIGURA 26 - Mapa temático do atributo espessura de parede categoria 3 (EPC_3)

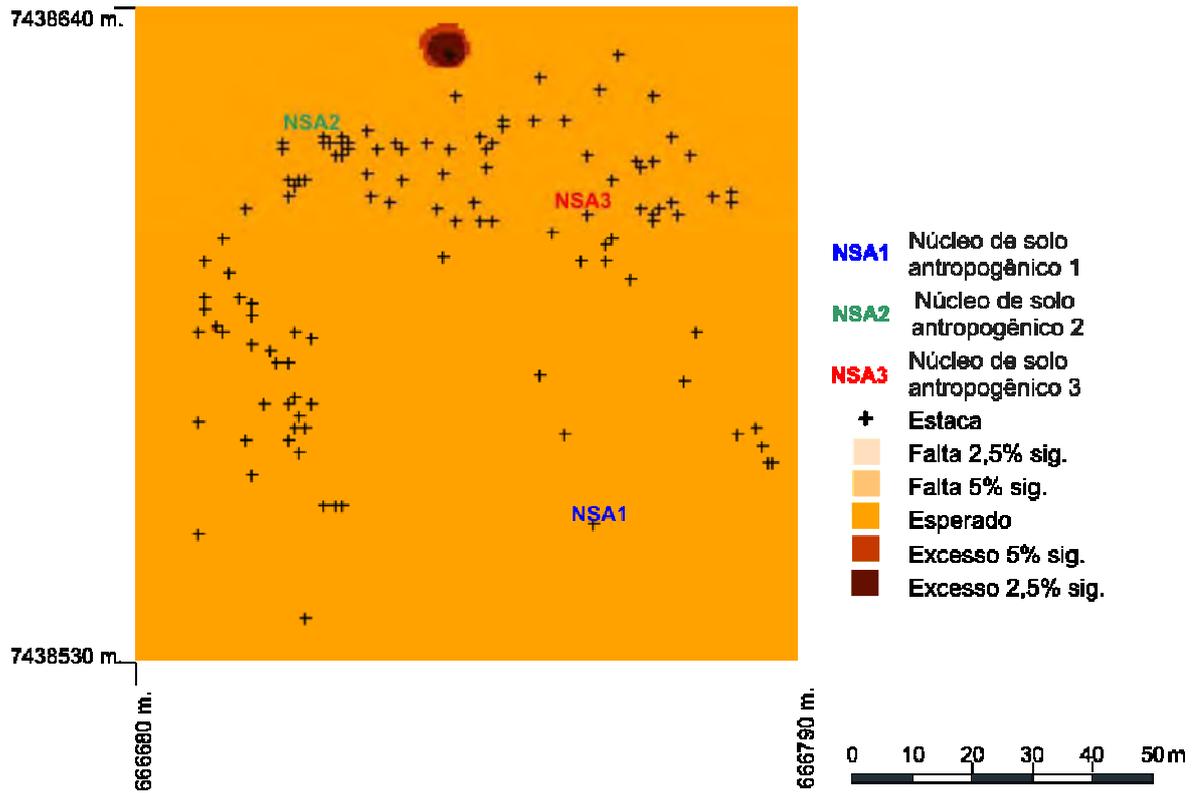


FIGURA 27 - Mapa temático do atributo queima categoria 3 (Q_3)

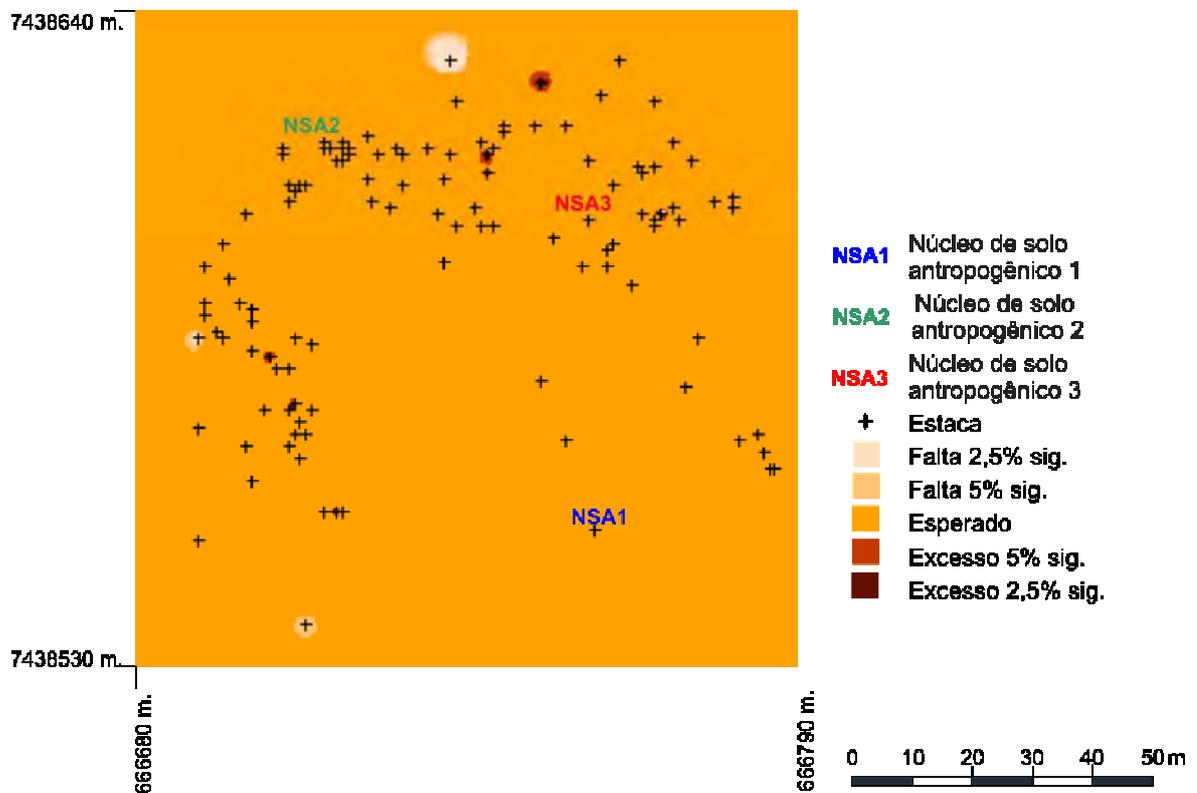


FIGURA 28 - Mapa temático do atributo queima categoria 5 (Q_5)

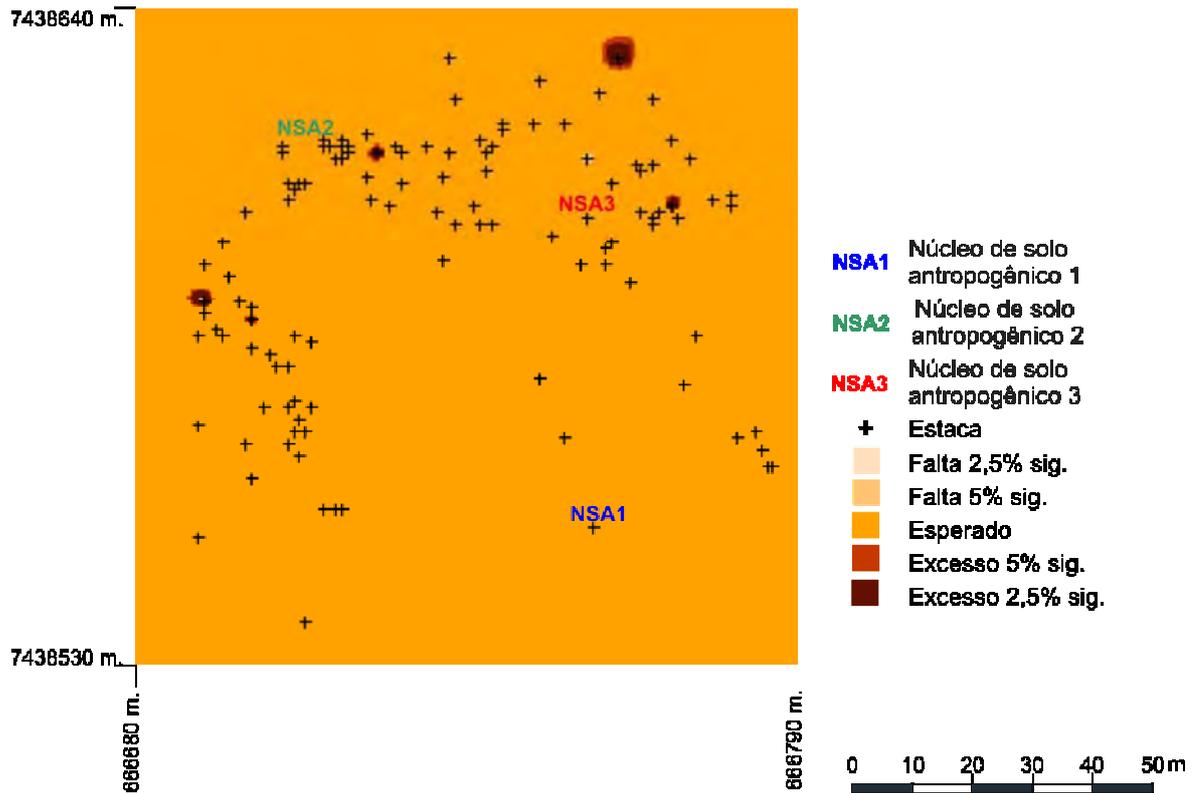


FIGURA 29 - Mapa temático do atributo queima categoria 6 (Q_6)

Estão associados entre si os atributos-categorias: CL_2, EPC_1, EPC_2 e DI_11, indicando que fragmentos cerâmicos de borda tendem a ter espessura de parede fina à média fina com decoração interna de engobo branco; DI_15, DE_15 e DE_11, indicando que fragmentos com decoração interna de engobo laranja tendem a ter decoração externa de engobo branco ou engobo laranja. Não foram detectadas associações relevantes dos atributos_categorias EPC_4 com outros atributos_categorias.

Considerando as restrições impostas pela técnica da análise de correspondência ao estudo conjunto de pontos linha e pontos coluna, essa análise privilegiará os pontos linhas/colunas que apresentam a maior contribuição à inércia total conjuntamente com a melhor qualidade de representação nas dimensões derivadas principais.

Segundo esse critério, a primeira análise a ser apresentada busca correlacionar o ponto linha NSA1 com os pontos coluna dos atributos_categorias nas dimensões onde se verifica a sua maior contribuição à inércia total, bem como a sua melhor qualidade de

representação nas dimensões derivadas 1 e 2.

A figura 30 ilustra o ponto linha NSA1, os pontos coluna representativos dos atributos_categorias, as dimensões derivadas auxiliares, bem como o tipo de relacionamento estabelecido entre o ponto linha NSA1 e os pontos coluna dos atributos_categorias.

Da análise da figura 30 pode-se dizer que o NSA1 estabelece correlação positiva com os atributos_categorias CMC_0-TP_1, CMC_2, CMC_3, MNC_2, MNC_3 e Q_4, correlação negativa com CMC_1 e MNC_1 e pouca ou nenhuma correlação com os demais atributos_categorias ditos comum. O comportamento espacial desses atributos_categorias pode ser melhor explorado nos mapas temáticos dos resíduos padronizados ilustrados nas figuras 31, 32, 33, 34, 35, 36, 37 e 38, respectivamente.

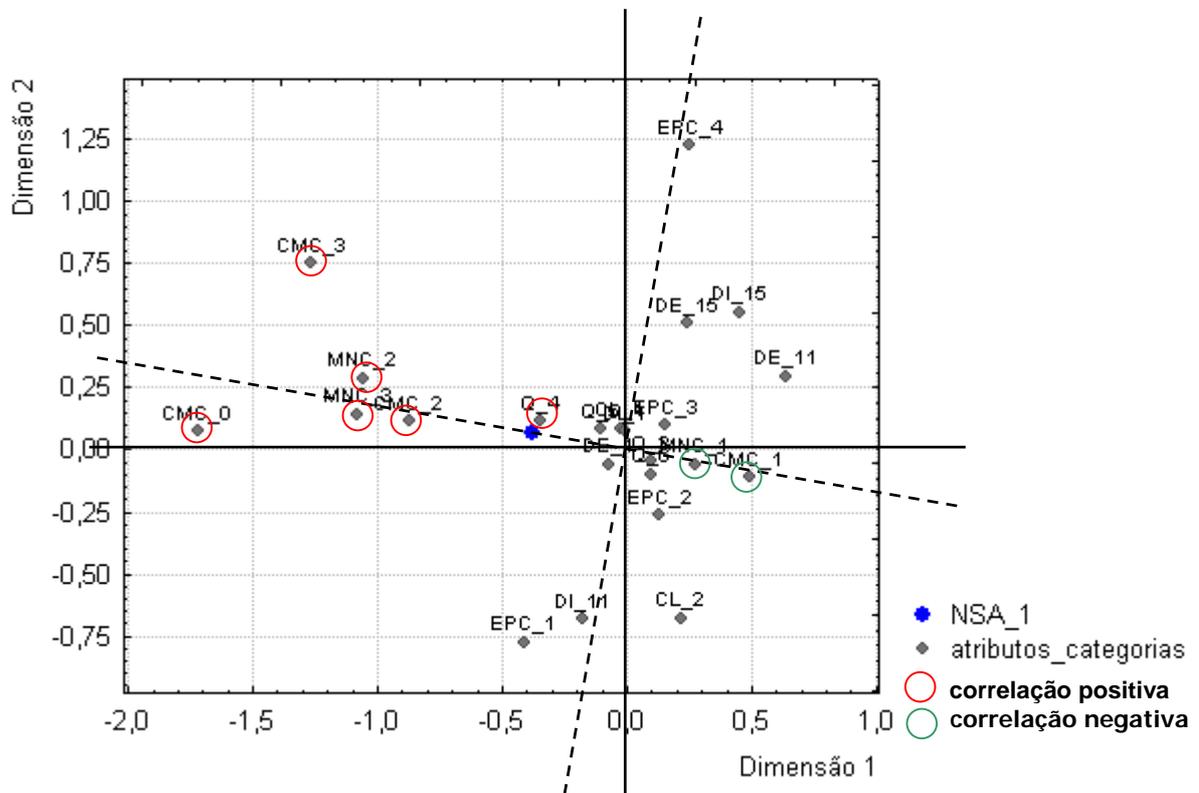


FIGURA 30 - Gráfico de dispersão das dimensões derivadas 1X2 do NSA1 conjunto com os atributos_categorias da tabela fcomum

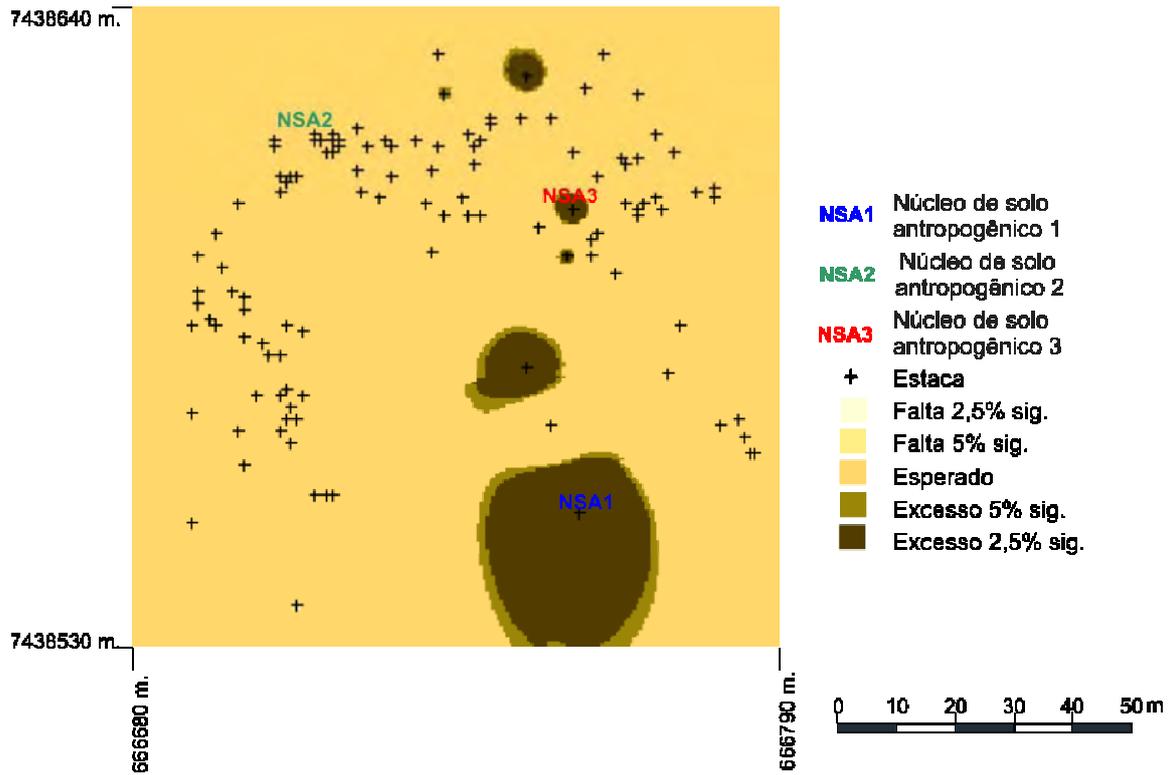


FIGURA 31 - Mapa temático do atributo caco moído categoria 0 (CMC_0) ou antiplástico categoria 1 (TP_1)

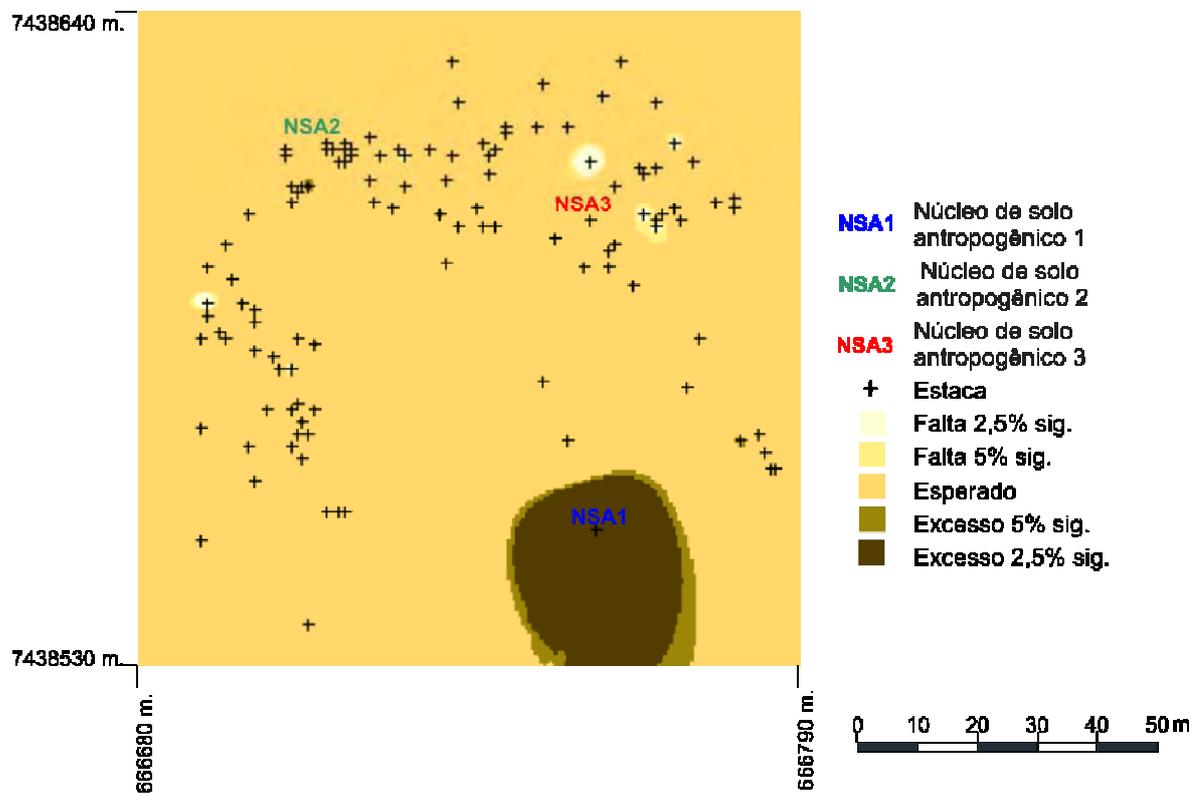


FIGURA 32 - Mapa temático do atributo caco moído categoria 2 (CMC_2).

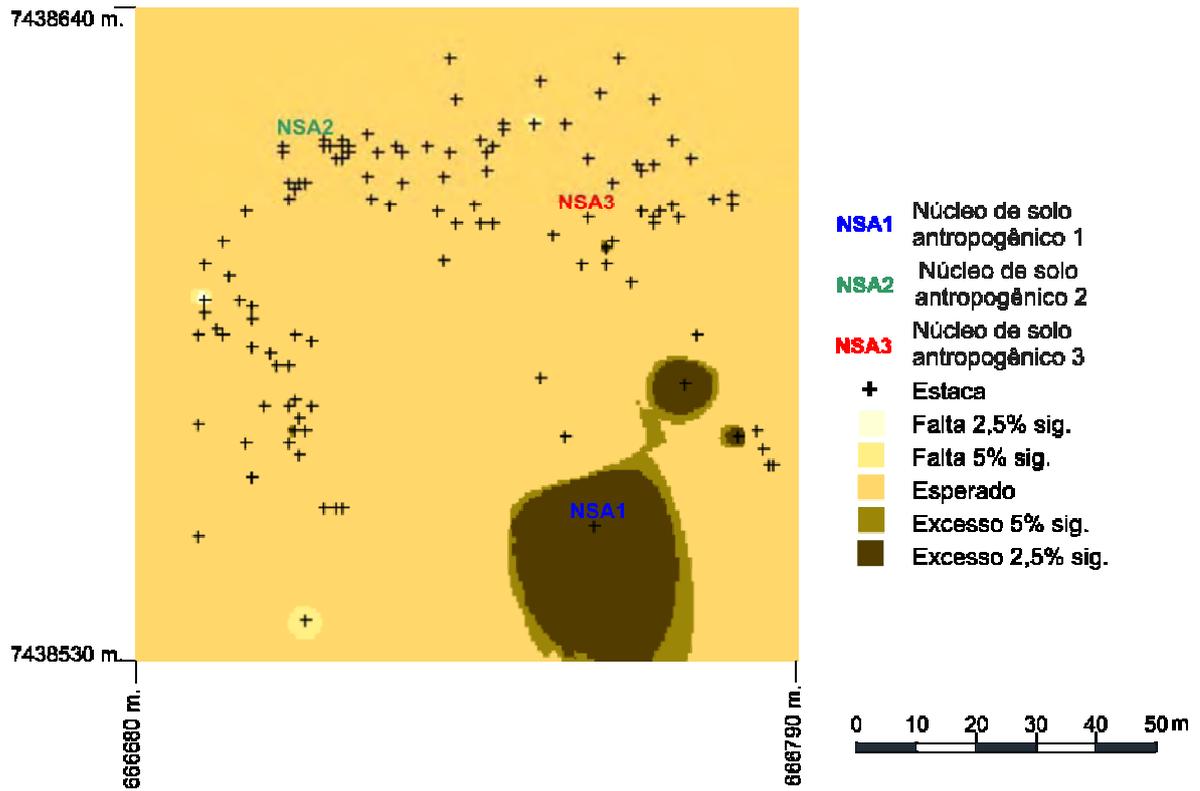


FIGURA 33 - Mapa temático do atributo caco moído categoria 3 (CMC_3).

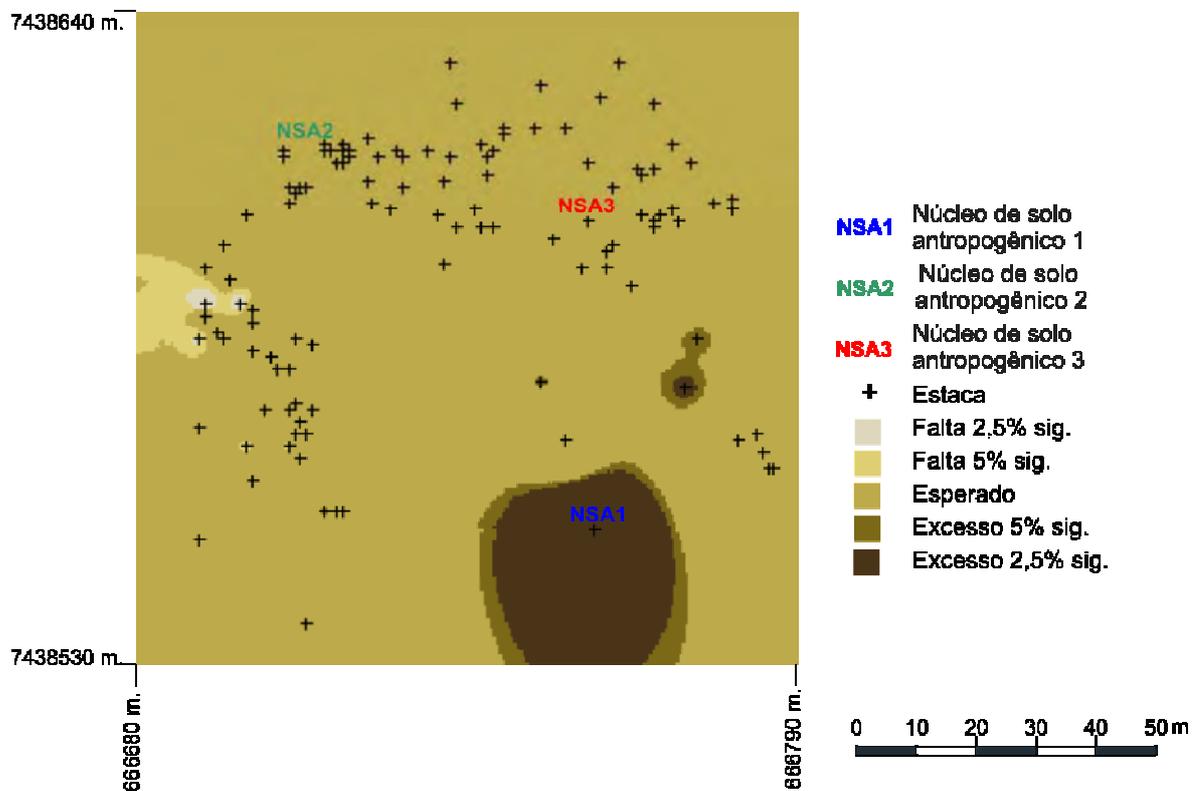


FIGURA 34 - Mapa temático do atributo material mineral categoria 2 (MNC_2)

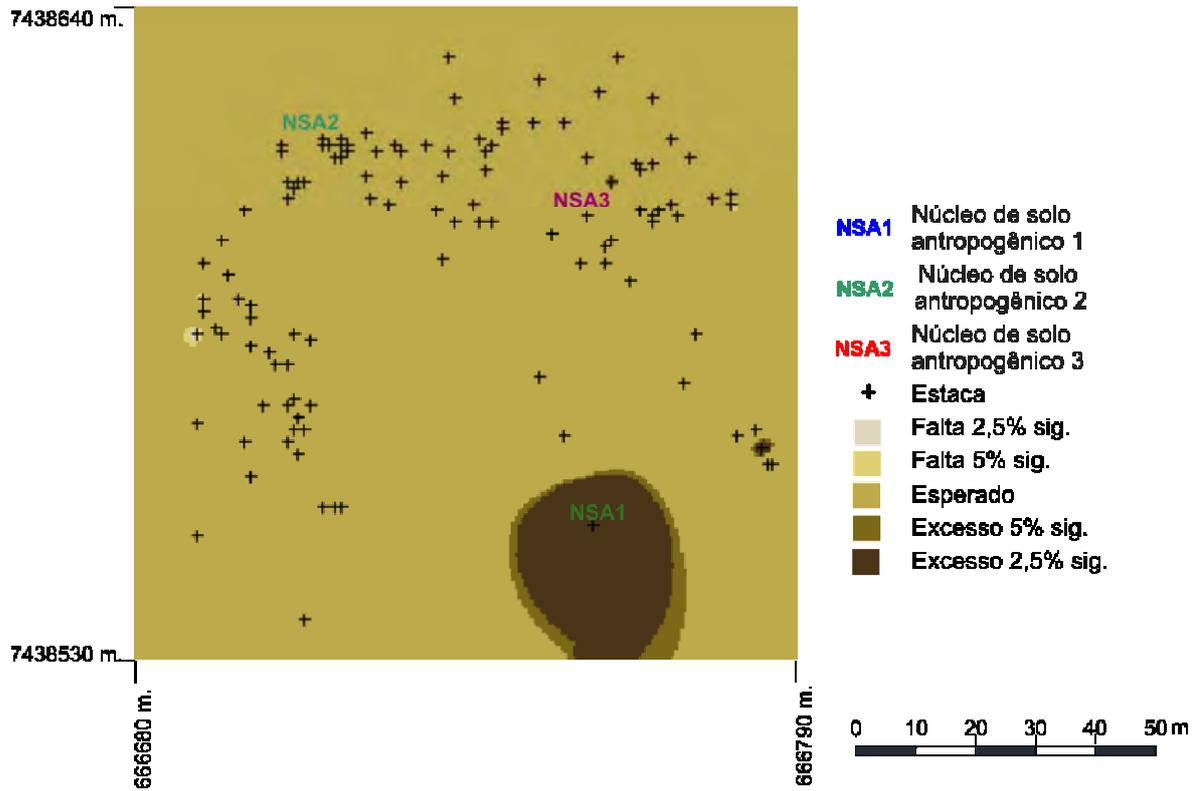


FIGURA 35 - Mapa temático do atributo material mineral categoria 3 (MNC_3)

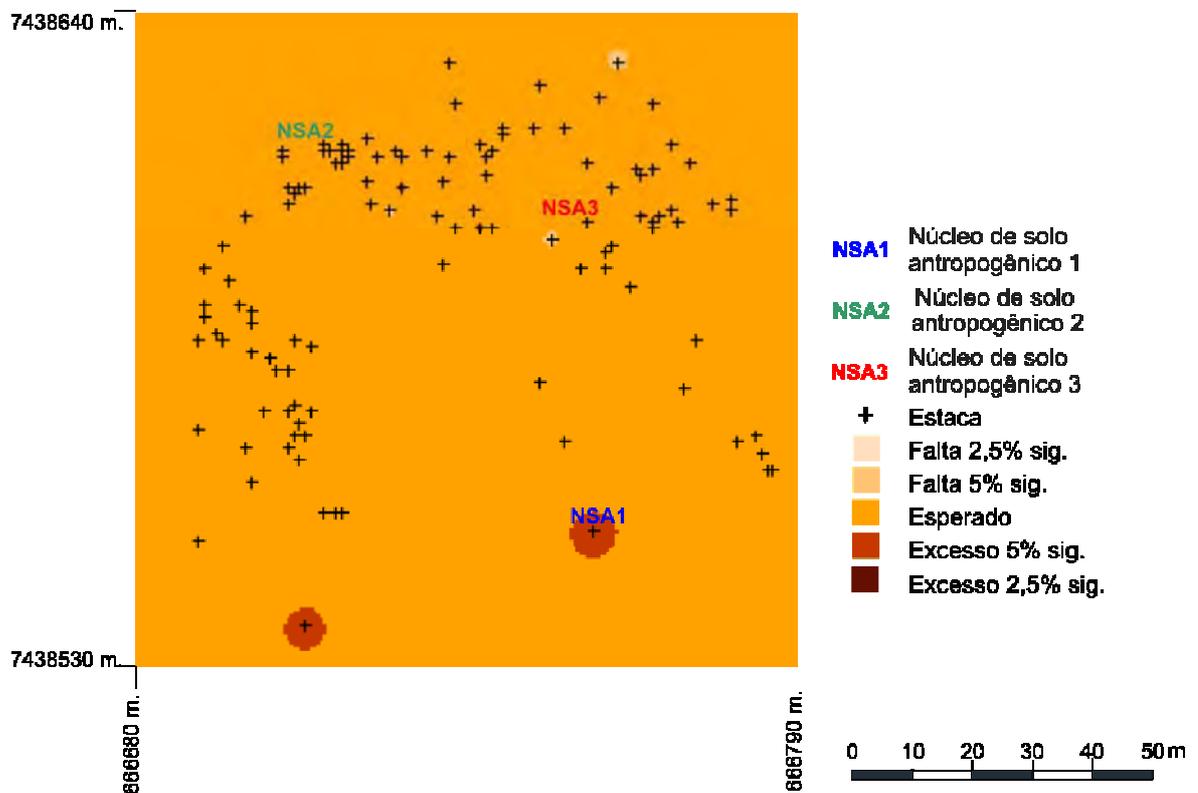


FIGURA 36 - Mapa temático do atributo queima categoria 4 (Q_4).

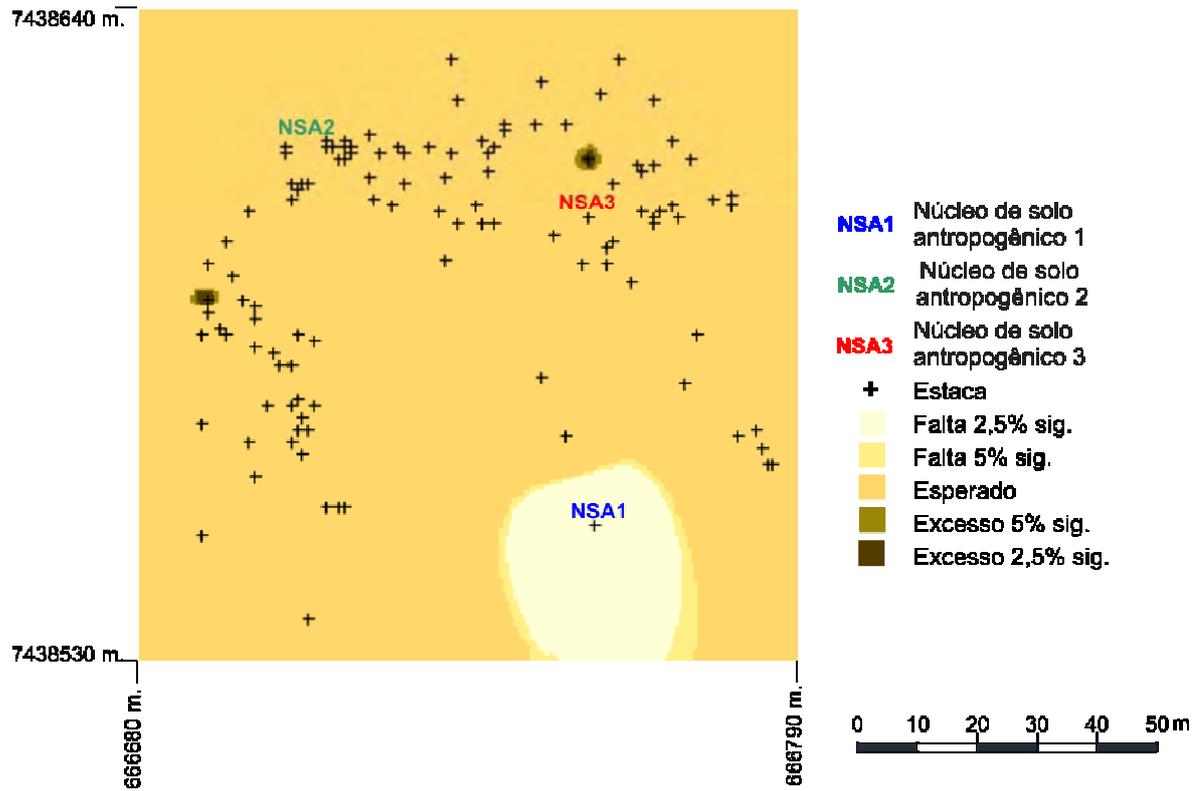


FIGURA 37 - Mapa temático do atributo caco moído categoria 1 (CMC_1).

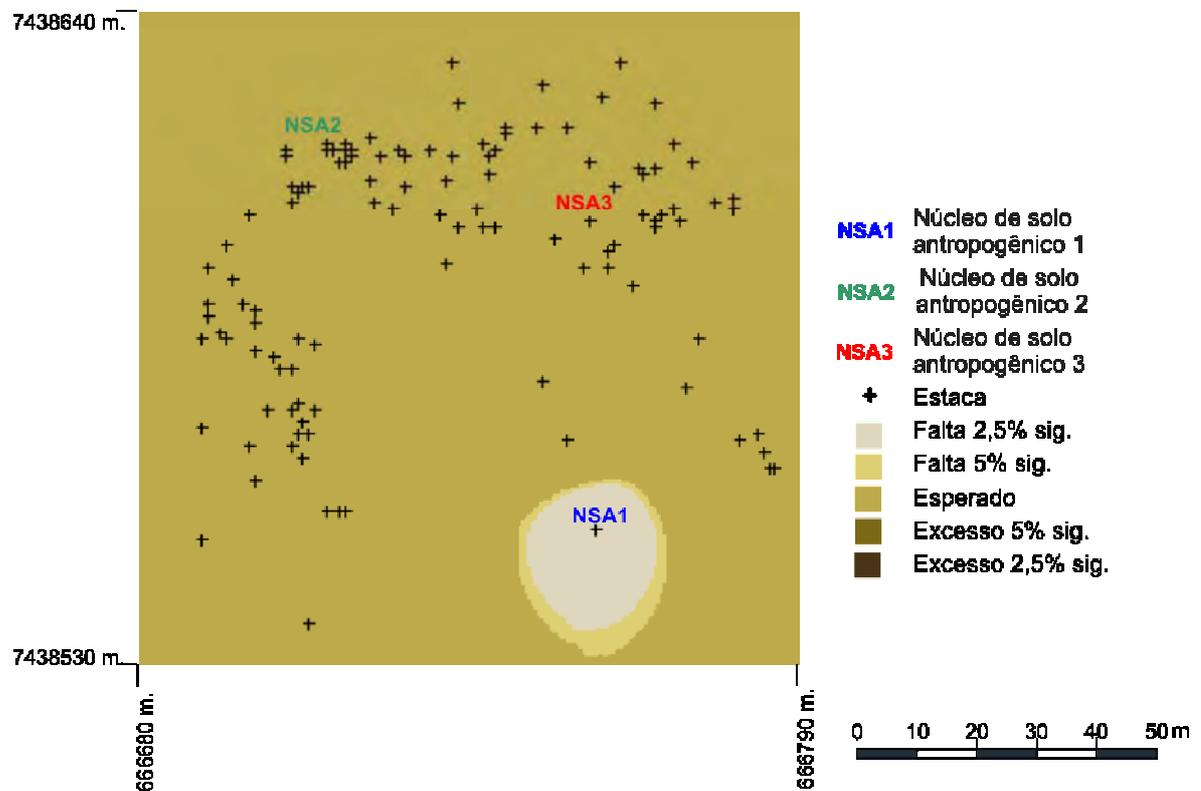


FIGURA 38 - Mapa temático do atributo material mineral categoria 1 (MNC_1).

Observa-se no NSA1 a concentração de fragmentos cujo material constituinte é o antiplástico do tipo 1 (material mineral) e do antiplástico do tipo 13 (material mineral + caco moído), onde o caco moído e o material mineral apresentam granulometria média e grossa em detrimento do caco moído e do material mineral de granulometria mais fina. Fica claro, então, pelo exame dos mapas temáticos, que o NSA1 se diferencia dentro do sítio arqueológico pelos atributos do material constituinte dos fragmentos cerâmicos ali coletados.

No NSA3 e algumas estacas em seu entorno foram observados pontos de concentração de fragmentos constituídos de antiplástico do tipo 1 (material mineral), e de antiplástico do tipo 13, existindo concentração do caco moído de granulometria fina em detrimento do caco moído de granulometria média.

Os atributos_categorias CL_2, EPC_1, EPC_4, DI_11, DI_15, DE_11 e DE_15 apresentam a sua melhor qualidade de representação nas dimensões derivadas 2 e 3.

As figuras 39 e 40 apresentam os gráficos de dispersão nas dimensões 2X3 desses pontos coluna atributos_categorias conjunto com os pontos linha dos NSAs e estacas, na qual o correlacionamento positivo/negativo entre os atributos_categorias e as estacas e NSAs é definido pela área de influência (em cor) dos atributos_categorias.

Da análise dos gráficos das figuras 39 e 40 conclui-se que:

- do grupo de estacas e estacas do NSA2 que estabelece correlação positiva com os atributos_categorias EPC_1 e DI_11, parte se correlaciona positivamente com CL_2, estando correlacionadas negativamente em partes com EPC_4 e DE_11, EPC_4, DE_11 e DI_15, e DI_15 e DE_15;
- o grupo que se correlaciona positivamente aos atributos_categorias DI_15 e DE_15 se correlaciona negativamente com CL_2;
- o grupo que se correlaciona positivamente com EPC_4 e DE_11 estabelece correlação negativa com EPC_1 e DI_11;
- o NSA3 estabelece correlação negativa com DE_11 e EPC_4.

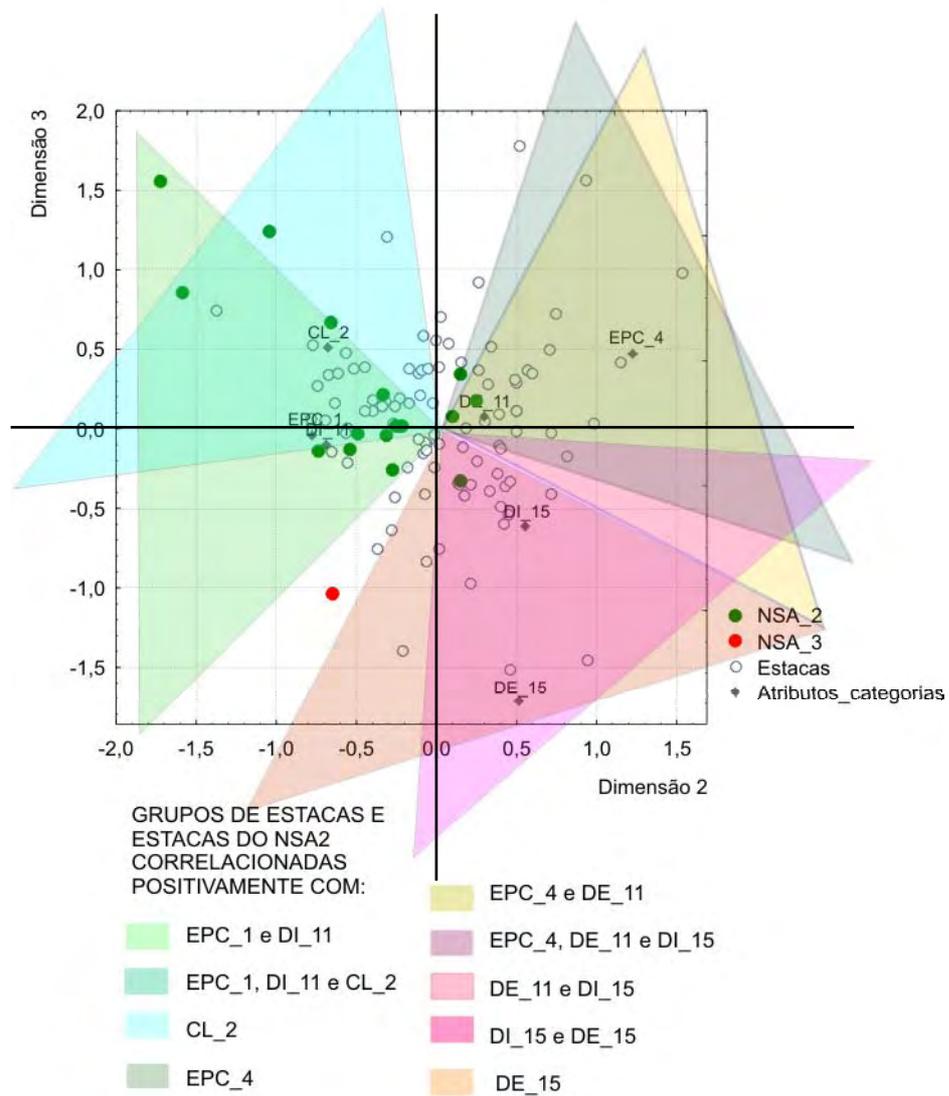


FIGURA 39 - Gráfico de dispersão das dimensões derivadas 2X3 das estacas, NSA2 e NSA3, conjunto com atributos_categorias da tabela fcomum, apresentando o correlacionamento positivo entre estacas e NSAs e atributos_categorias.

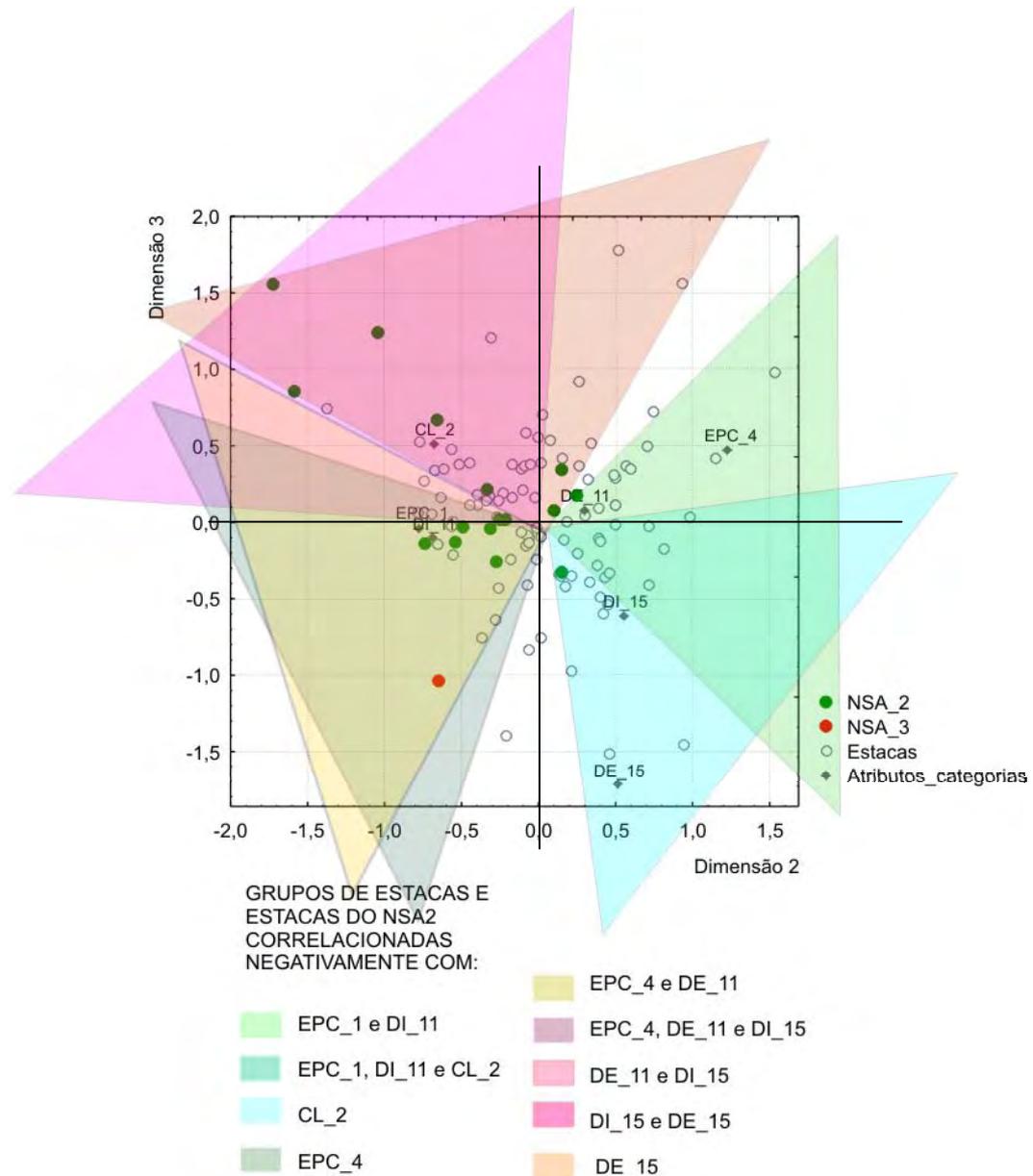


FIGURA 40 - Gráfico de dispersão das dimensões derivadas 2X3 das estacas, NSA2 e NSA3, conjunto com atributos_categorias da tabela fcomum, apresentando o correlacionamento negativo entre estacas e NSAs e atributos_categorias.

Os mapas temáticos dos resíduos padronizados dos atributos EPC_1 e DI_11, correlacionados ao mesmo grupo de estacas, e CL_2, que se correlaciona com parte desse grupo de estacas, ilustram as figuras 41, 42 e 43, respectivamente, e apresentam pontos de concentração/falta desses atributos esparsos em todo o Sítio Piracanjuba. O fato dos pontos não estabelecerem vizinhança ou proximidade entre si, impossibilita a sua vinculação a áreas específicas do sítio, portanto o estudo desses atributos_categorias só é possível ponto a ponto,

ou seja, estaca a estaca.

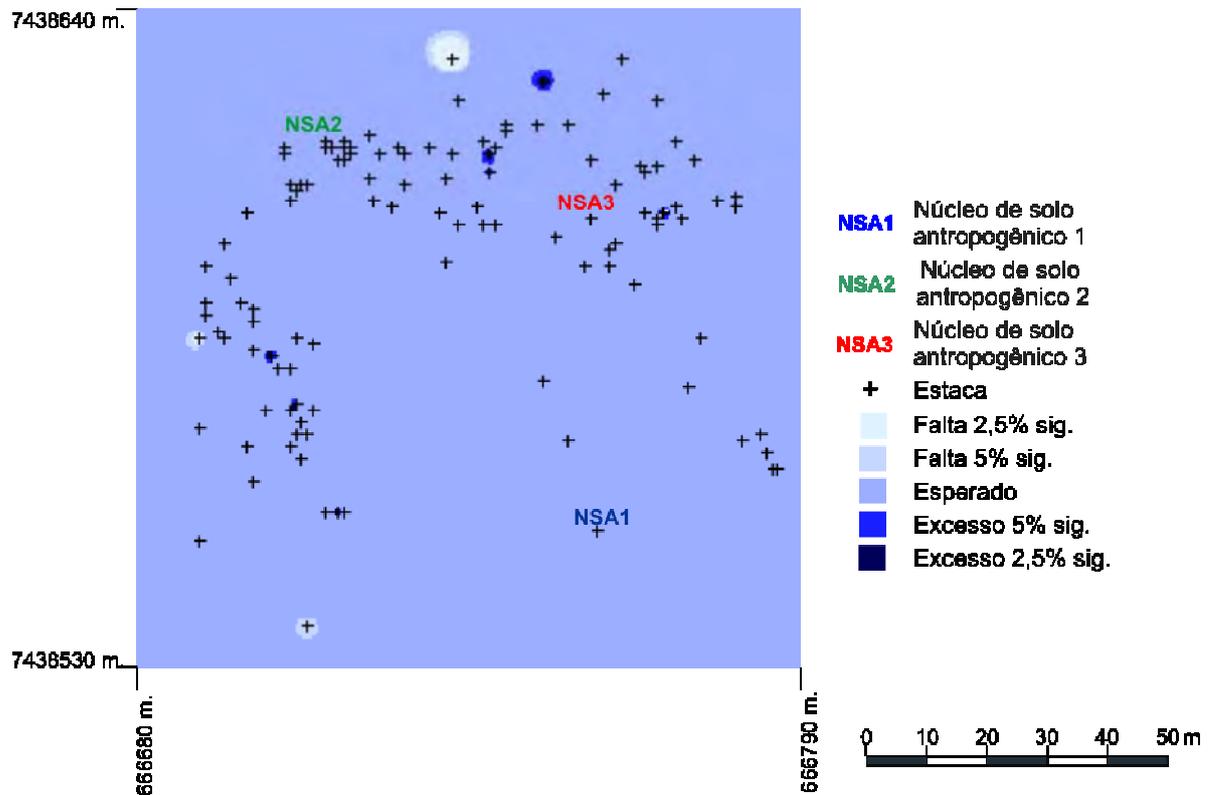


FIGURA 41 - Mapa temático do atributo espessura de parede categoria 1 (EPC_1).

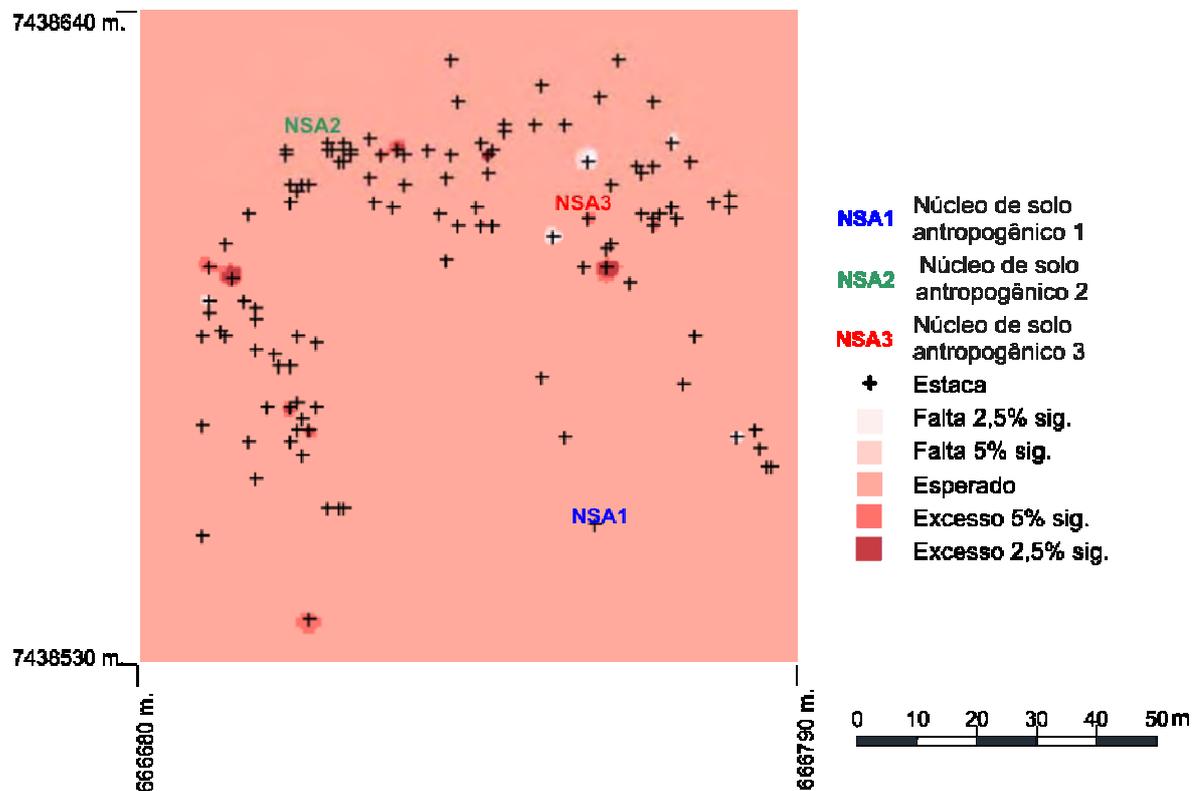


FIGURA 42 - Mapa temático do atributo decoração interna categoria 11 (DI_11).

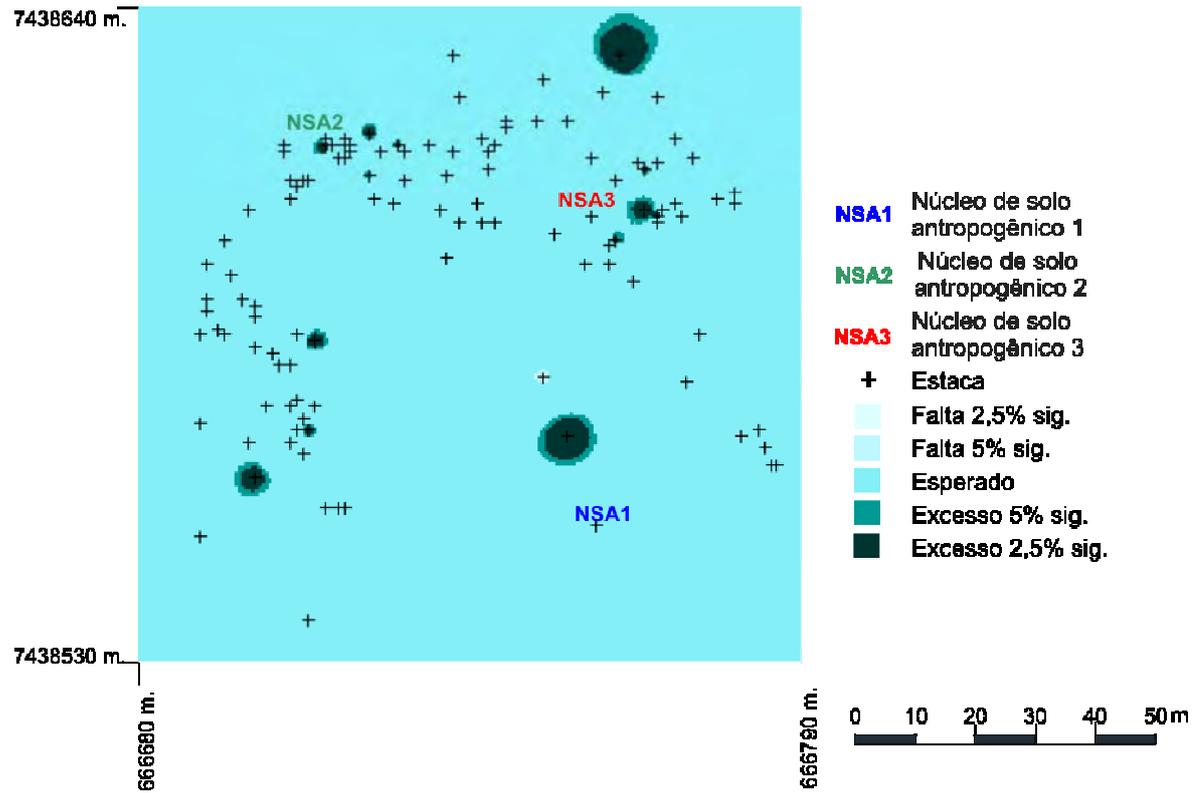


FIGURA 43 - Mapa temático do atributo classe categoria 2 (CL_2).

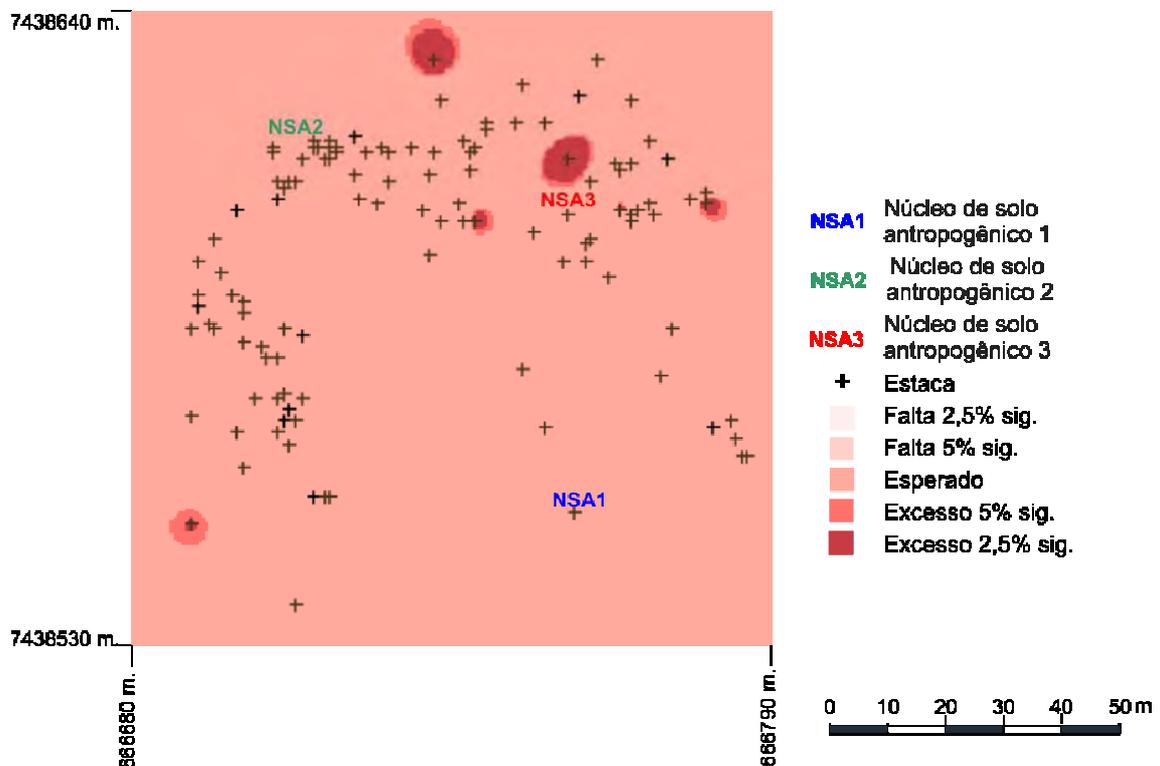


FIGURA 44 - Mapa temático do atributo decoração interna categoria 15 (DI_15).

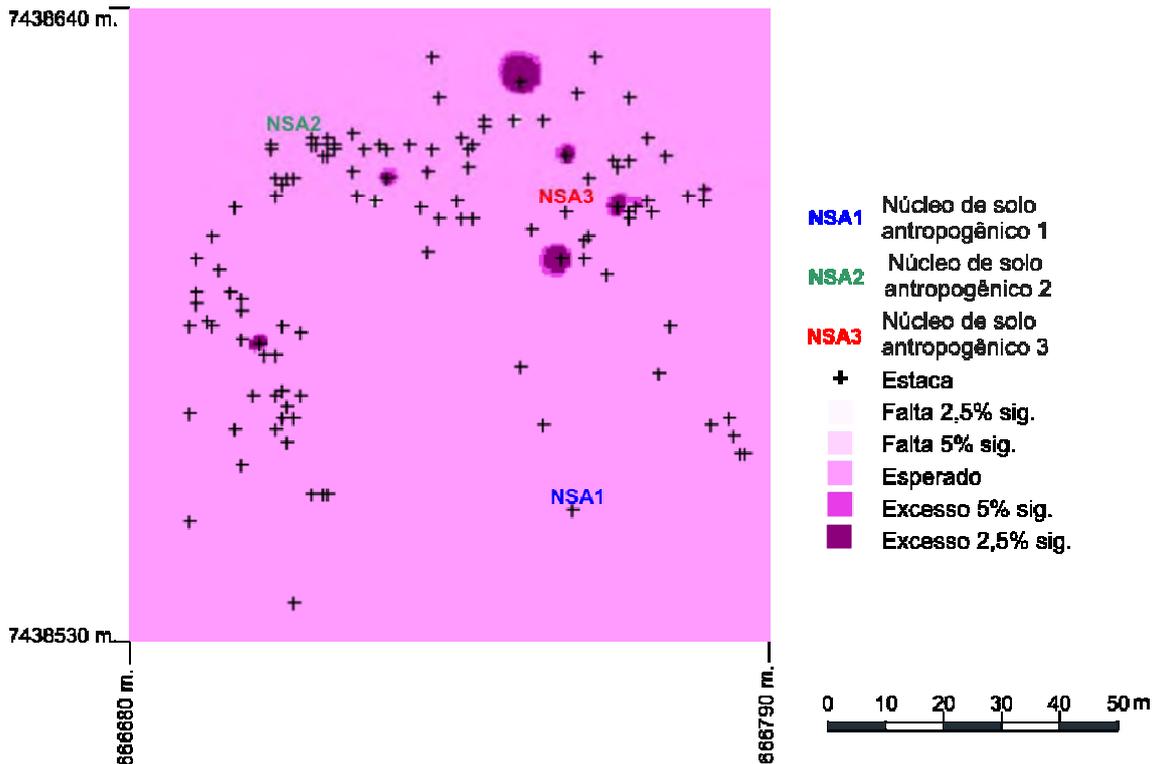


FIGURA 45 - Mapa temático do atributo decoração externa categoria 15 (DE_15).

Nos mapas temáticos dos resíduos padronizados dos atributos_categorias DI_15 e DE_15 apresentados nas figuras 44 e 45, respectivamente, observa-se a relação espacial dos mesmos caracterizada pela concentração de fragmentos cerâmicos com esses atributos_categorias em estacas esparsas e em estacas no entorno do NSA3.

Os mapas temáticos dos resíduos padronizados dos atributos_categorias DE_11 e EPC_4, este último mais fortemente vinculado a determinados grupos de estacas posto que apresenta o maior afastamento do centro de distribuição de massa, ilustram as figuras 46 e 47, respectivamente, onde se observam pontos de concentração no entorno do NSA3.

Portanto, em termos dos atributos_categorias de fabricação ditos comuns há que se destacar a área no entorno do NSA3 como de concentração de fragmentos cerâmicos com espessuras de parede média grossa e grossa, fragmentos de borda, decoração interna de engobo branco ou laranja e decoração externa de engobo branco ou laranja; e a falta de fragmentos com decoração externa de engobo branco no NSA1.

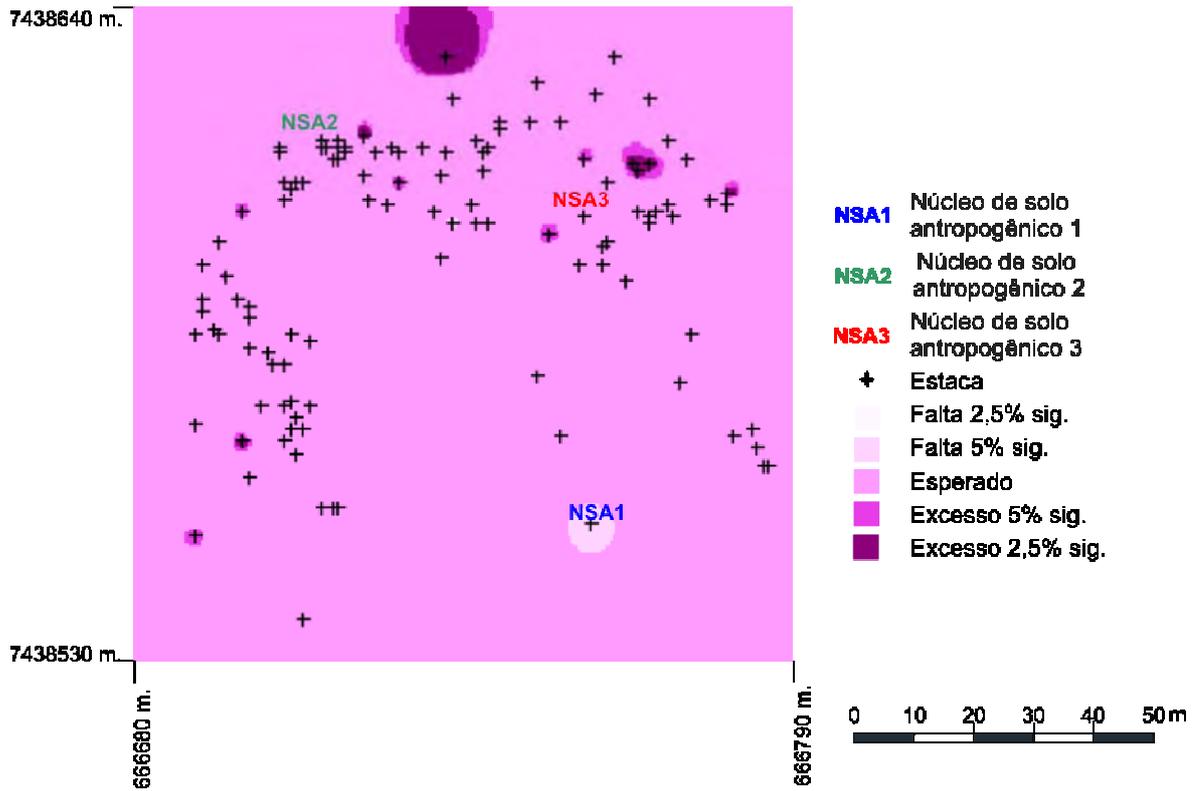


FIGURA 46 - Mapa temático do atributo decoração externa categoria 11 (DE_11)

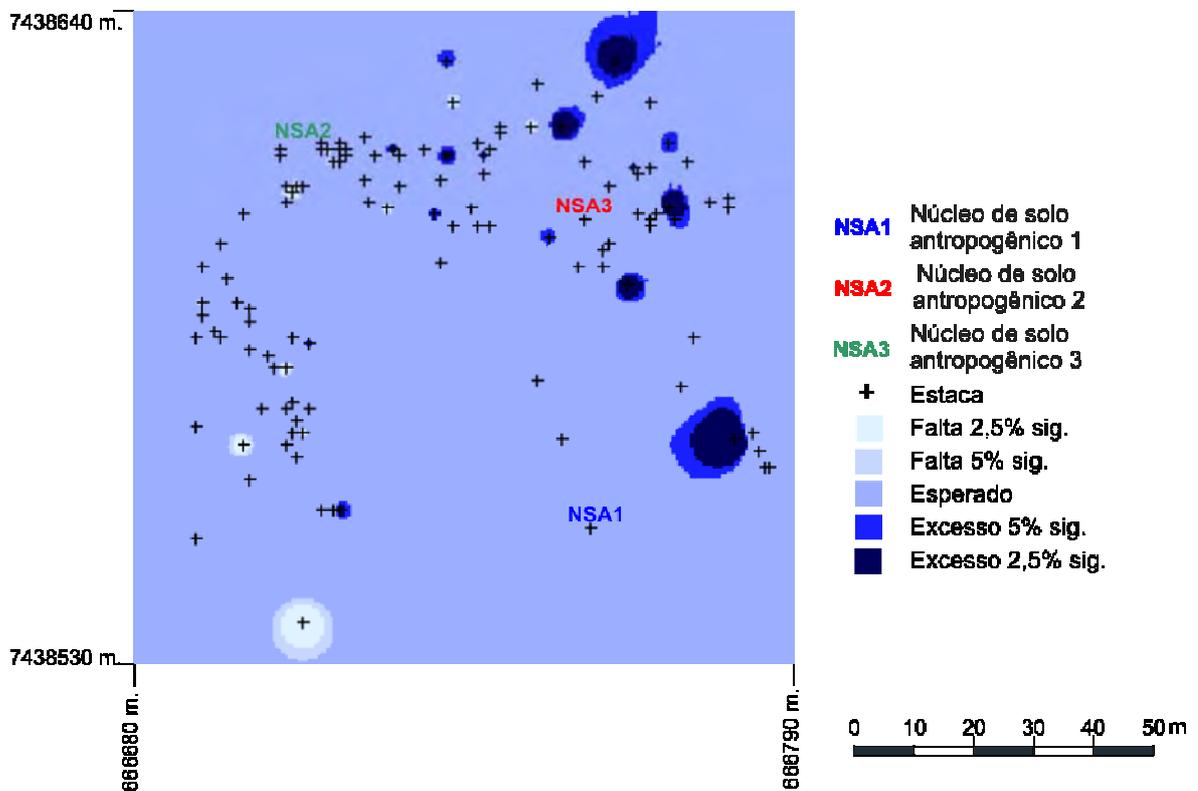


FIGURA 47 - Mapa temático do atributo espessura de parede categoria 4 (EPC_4).

1.13.1.2 Tabela de contingência raros

Quanto à tabela raros todos os fragmentos cerâmicos analisados nesta fase foram confeccionados segundo a mesma técnica de confecção, roletado, estando 96% em bom estado de conservação e 4% pouco conservados.

O gráfico de linhas dos perfis linha e do perfil marginal das linhas, que consta do apêndice A (mídia digital) no arquivo perfis – planilha perfis linha raros, apresenta grande dispersão em torno do perfil marginal das linhas, no qual se observa a exclusividade de alguns atributos_categorias em determinadas estacas e NSAs, por exemplo, todos os fragmentos considerados raros coletados na estaca 50 pertencem à mesma categoria de queima, categoria 1. A ocorrência de atributos_categorias exclusivos é pontual, não se registrando atributos_categorias dominantes.

O gráfico de linhas dos perfis coluna e do perfil marginal das colunas, que consta do arquivo perfis – planilha perfis coluna raros do apêndice A (mídia digital), apresenta o NSA1 como o ponto de maior coleta de fragmentos raros. O que não equivale a dizer que no NSA1 encontram-se concentrados fragmentos raros, pois como o NSA1 é o ponto onde houve a maior coleta de fragmentos cerâmicos seria de se esperar que a sua contribuição estivesse refletida em termos de fragmentos comuns, raros, raríssimos e excepcionais.

O perfil do NSA1 apresentado no gráfico de linhas foi substituído por um perfil simulado com valores correspondentes a 10% dos valores originais.

Isto posto, a dispersão em torno do perfil marginal indica que os fragmentos coletados possuem atributos_categorias vinculados aos pontos de coleta.

A inércia total do sistema é igual a 0,88920, indicando heterogeneidade na distribuição dos fragmentos cerâmicos com atributos_categorias considerados raros nas estacas e NSAs.

O valor do teste do χ^2 é igual a 1435,2 sendo significativo a 2,5% de significância (o valor crítico da distribuição χ^2 com 1040 graus de liberdade é 904,74, $p = 0,00$), isto é, existem associações/relações entre as linhas e colunas, entre os elementos das linhas e entre os elementos das colunas passíveis de serem estudados com a aplicação da técnica da análise de correspondência com normalização canônica.

A tabela 2 apresenta os valores singulares, os autovalores, a inércia explicada a cada dimensão e a inércia explicada acumulada. As três primeiras dimensões contribuem com 50,0983% (em negrito) da inércia total.

O exame da figura 48 que apresenta o gráfico dos autovalores de cada dimensão derivada, não auxilia na determinação do número de dimensões derivadas capaz de representar adequadamente o sistema sob análise.

Isto posto, a análise das tabelas eletrônicas linhas raras e colunas raras do arquivo dimensões (apêndice A – mídia digital) mostram que as linhas/colunas que mais contribuem à inércia total do sistema apresentam boa qualidade de representação nas três primeiras dimensões derivadas, justificando desta forma a escolha das três primeiras dimensões derivadas para este estudo.

Tabela 2 - Valores singulares, autovalores e inércia – tabela fraros.

Dimensão	Valores Singulares	Autovalores	Inércia Explicada a cada dimensão	Inércia explicada acumulada
1	0,411498	0,169331	19,04315	19,0432
2	0,383488	0,147063	16,53885	35,5820
3	0,359275	0,129079	14,51634	50,0983
4	0,319600	0,102144	11,48728	61,5856
5	0,301333	0,090802	10,21167	71,7973
6	0,269933	0,072864	8,19436	79,9917
7	0,252866	0,063941	7,19094	87,1826
8	0,231034	0,053377	6,00283	93,1854
9	0,185490	0,034407	3,86941	97,0548
10	0,161828	0,026188	2,94517	100,0000

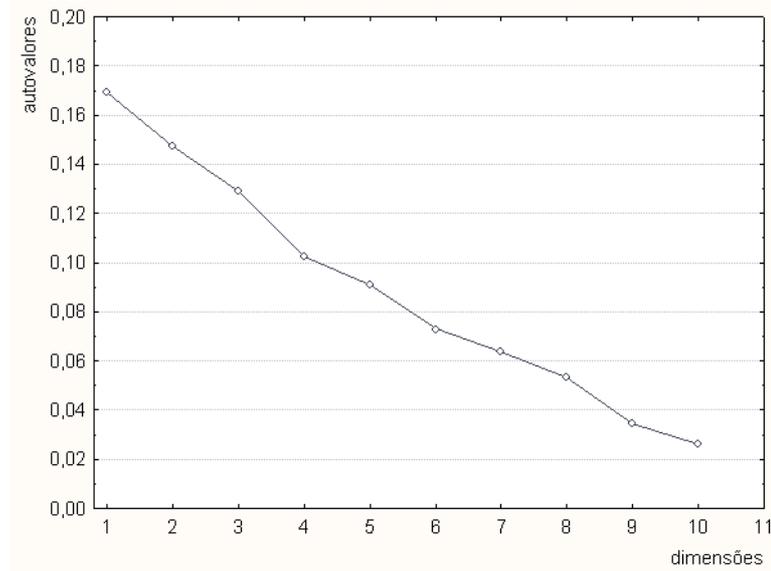


FIGURA 48 - Gráfico das dimensões derivadas – tabela fraros

Os gráficos de dispersão das dimensões derivadas das estacas e NSAs da tabela fraros ilustrados nas figuras 49, 50 e 51, dimensões 1X2, 1X3 e 2X3, respectivamente, mostram que o NSA2 se encontra disperso ao redor da origem das dimensões derivadas, apresentando estacas que se aproximam da origem, indicando certa homogeneidade proporcional na maneira como os atributos_categorias se distribuem nas mesmas e algumas estacas que se afastam muito do centro de distribuição de massa, configurando-se como “outliers”, indicando a concentração de atributos_categorias.

As estacas encontram-se bastante dispersas ao longo dos eixos derivados 2 e 3, indicando a vinculação de atributos_categorias a determinados grupos de estacas e apresentando elementos que se configuram como “outliers”.

O NSA1 apresenta vinculação ao primeiro eixo e terceiro eixo derivado e o NSA3 ao segundo e terceiro eixos derivados.

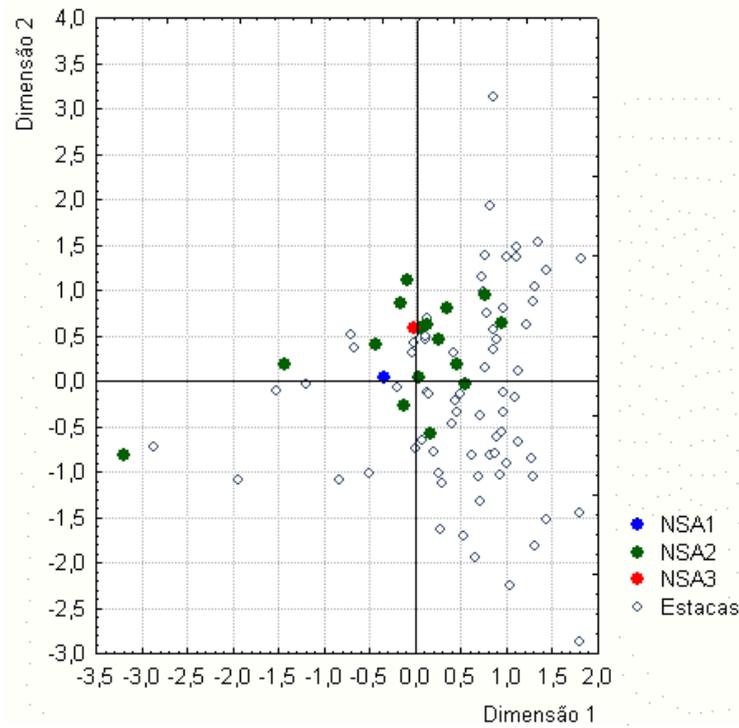


FIGURA 49 - Gráfico de dispersão das dimensões derivadas 1X2 das estacas e NSAs da tabela fraros

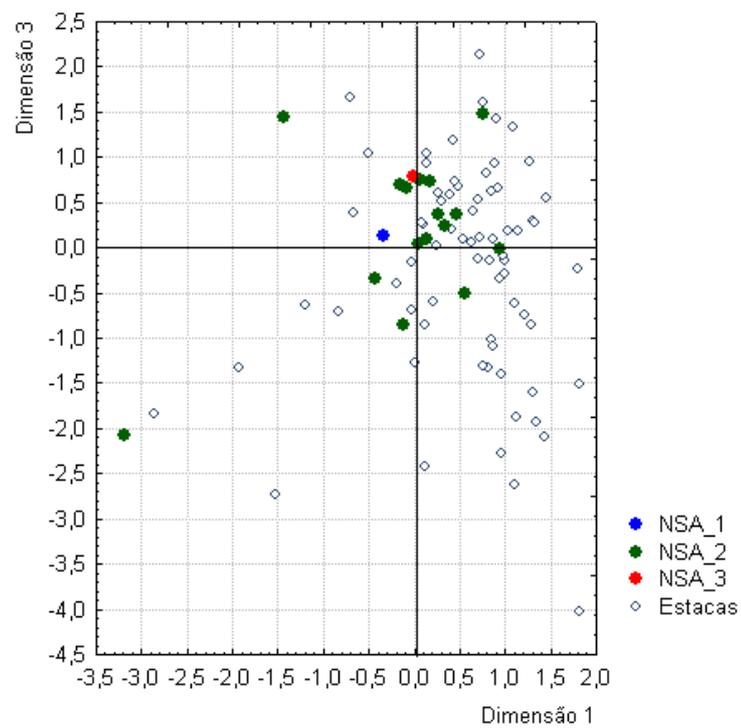


FIGURA 50 - Gráfico de dispersão das dimensões derivadas 1X3 das estacas e NSAs da tabela fraros

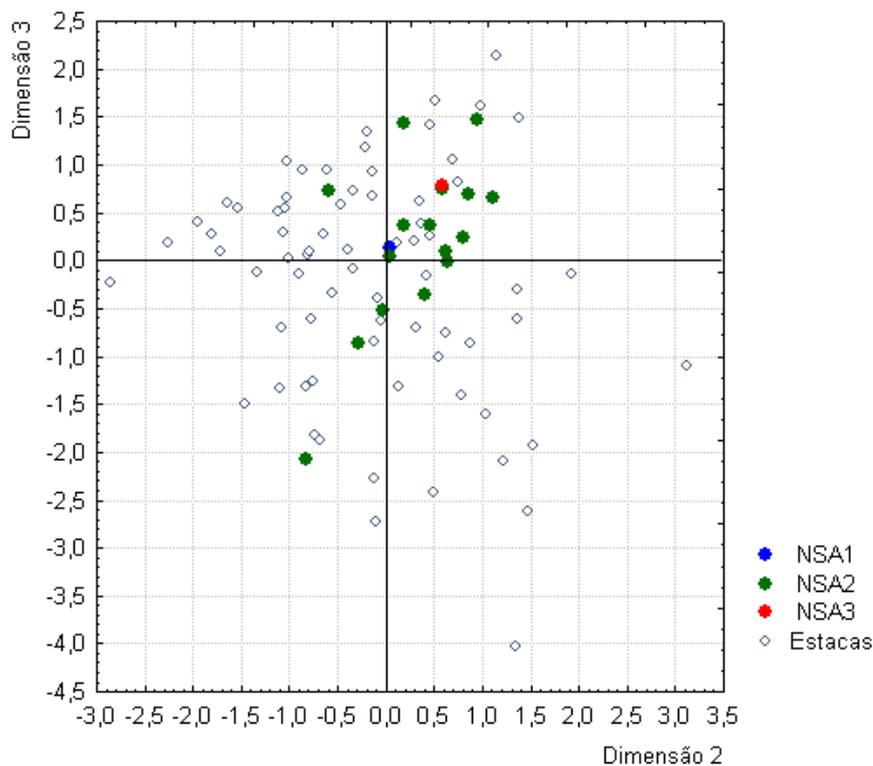


FIGURA 51 - Gráfico de dispersão das dimensões derivadas 2X3 das estacas e NSAs da tabela fraros

No exame dos gráficos de dispersão das dimensões derivadas dos atributos_categorias em suas dimensões derivadas 1X2, 1X3 e 2X3 apresentados nas figuras 52, 53 e 54, respectivamente, há que se observar a contribuição à inércia total e a qualidade da representação dos pontos atributos_categorias (tabela eletrônica colunas raros do arquivo dimensões (apêndice A – média digital)).

Os atributos_categorias Q_0 e Q_1 apresentam baixa contribuição à inércia total, bem como má qualidade de representação nas 3 dimensões derivadas eleitas para esse estudo, justificando a sua desincorporação do estudo da associação/relação entre atributos_categorias, bem como da associação/relação estacas e NSAs e atributos categorias. Esse fato não impede o estudo isolado dos mapas temáticos dos resíduos padronizados desses atributos_categorias apresentados nas figuras 55 e 56, respectivamente.

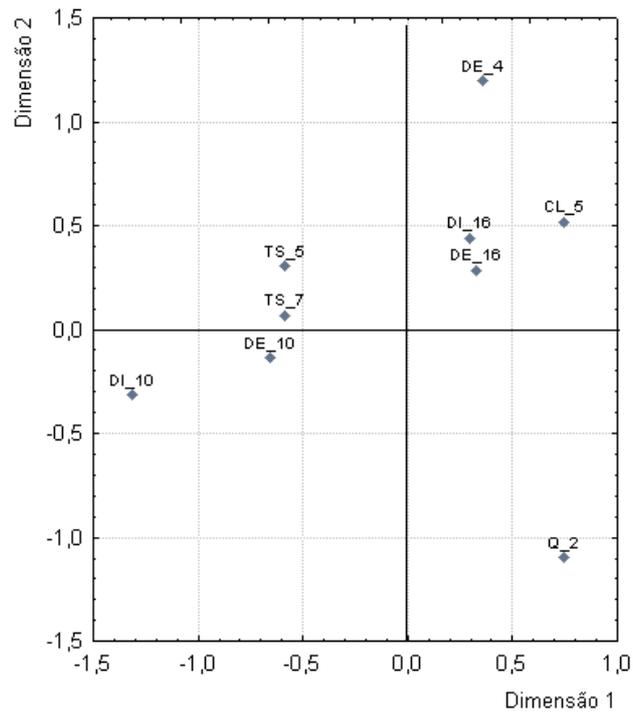


FIGURA 52 - Gráfico de dispersão das dimensões derivadas 1X2 dos atributos_categoriais da tabela fraros

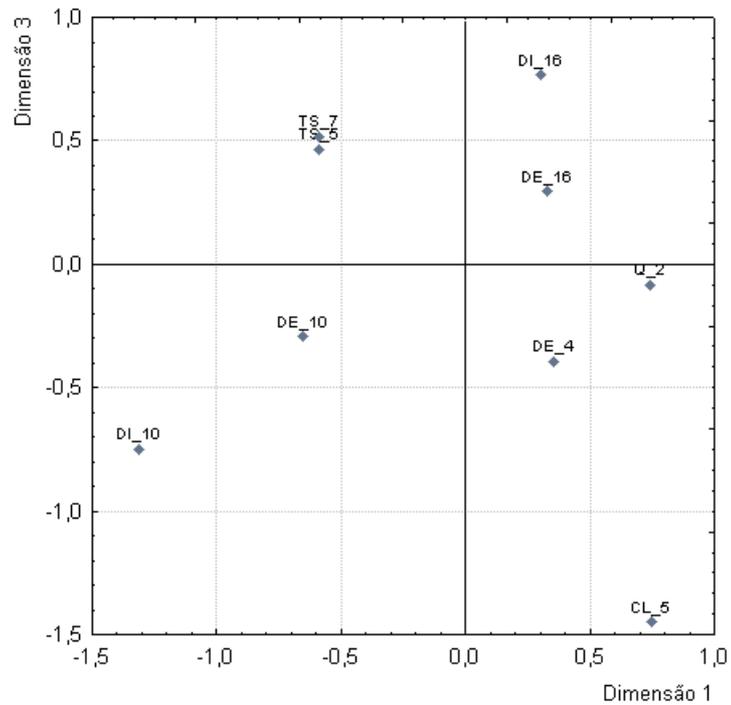


FIGURA 53 - Gráfico de dispersão das dimensões derivadas 1X3 dos atributos_categoriais da tabela fraros

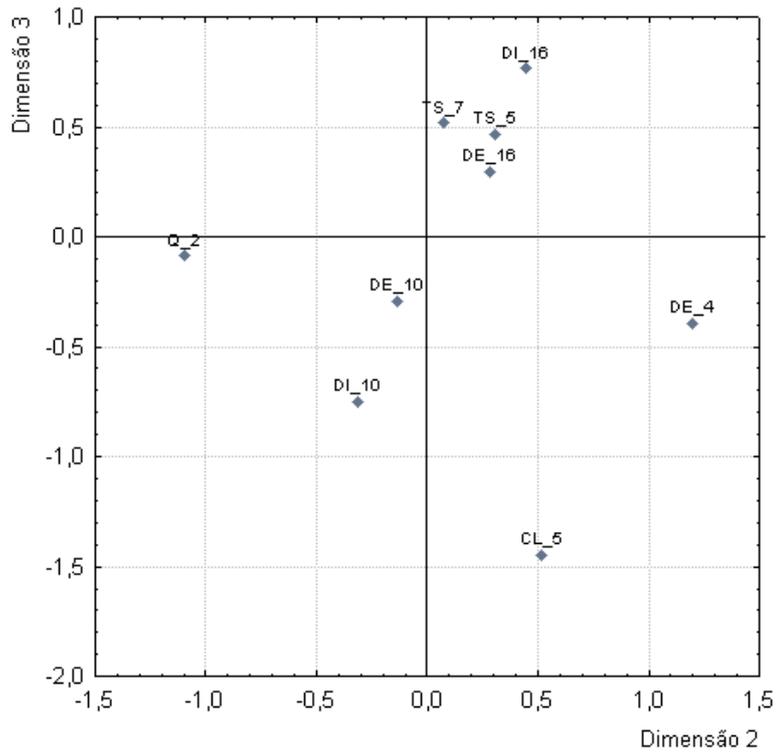


FIGURA 54 - Gráfico de dispersão das dimensões derivadas 2X3 dos atributos_categorias da tabela fraros

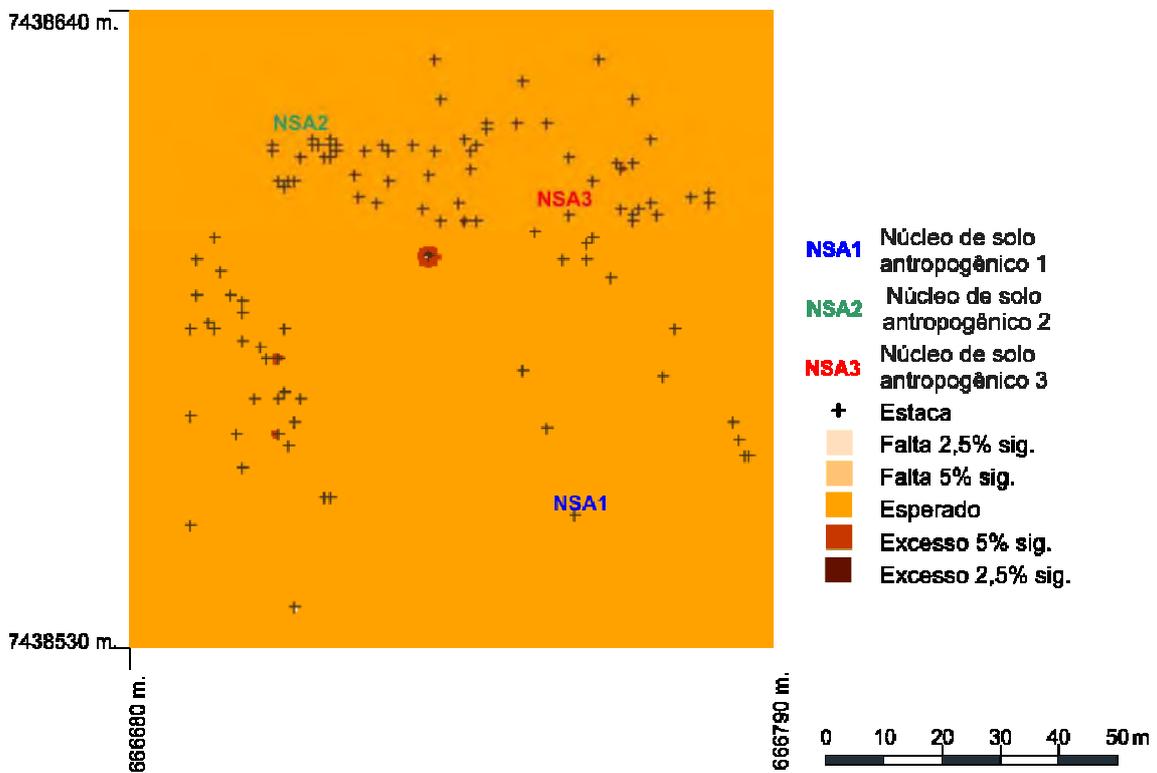


FIGURA 55 - Mapa temático do atributo queima categoria 0 (Q_0).

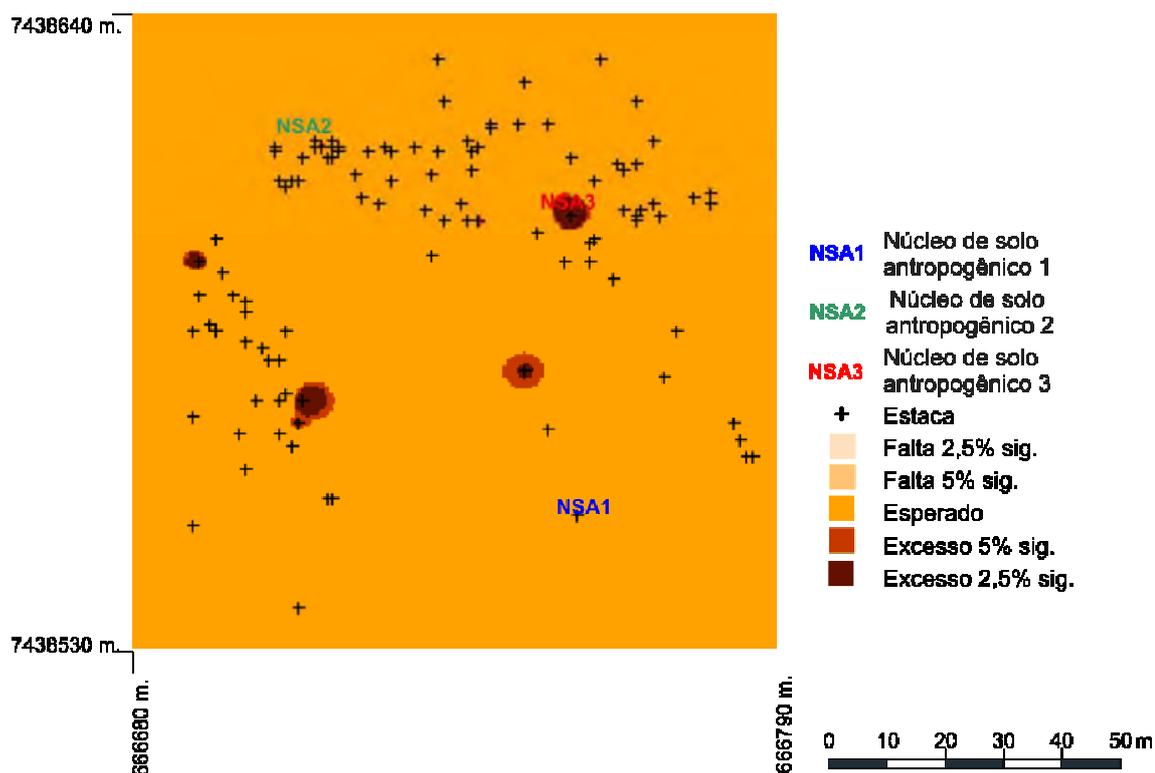


FIGURA 56 - Mapa temático do atributo queima categoria 1 (Q_1).

A análise conjunta dos gráficos 52, 53 e 54 revela que:

- os atributos TS_5 e TS_7 encontram-se associados entre si, o que em princípio poderia parecer uma incongruência, mas a descrição da categoria 5 e 7 desses atributos justifica a associação. A categoria 5 corresponde ao tratamento superficial polimento interno/alisamento interno e a 7 a polimento interno/externo, ou seja, as categorias 5 e 7 não são excludentes;
- os atributos DI_10 e DE_10 encontram-se associados o que significa que fragmentos cerâmicos com decoração interna de engobo vermelho tendem a ter decoração externa de engobo vermelho;
- a associação entre DI_16 e DE_16 indica que fragmentos com decoração interna pintado tendem a ter decoração externa pintado;

- a posição e o afastamento do centro de distribuição de massa dos atributos_categorias CL_5, Q_2, DI_10 e DE_4, indicam forte concentração e vinculação dos mesmos a grupos de estacas distintas.
- os demais atributos_categorias não estabelecem associações/relações entre si.

Na análise dos mapas temáticos dos resíduos padronizados dos atributos_categorias Q_0 e Q_1, figuras 55 e 56, há que se destacar a concentração de fragmentos cerâmicos com Q_1 no NSA3.

As figuras 57 e 58 apresentam as estacas e NSAs conjuntamente com os atributos_categorias e a correlação que estabelecem entre si.

Na análise conjunta das estacas e NSAs e dos atributos_categorias da tabela fraros proceder-se-á primeiramente a do NSA1, que dentre os pontos linha apresenta a maior contribuição à inércia total.

Dos gráficos 57 e 58, o NSA1 e um pequeno conjunto de estacas se encontram fortemente correlacionados positivamente com os atributos categorias TS_5 e TS_7, associados entre si com DE_10 e DI_10, negativamente com DE_16, DI_16, CL_5 e Q_2, não se correlacionando com DE_4.

Os mapas temáticos dos resíduos padronizados dos atributos_categorias que mais se correlacionam positivamente com o NSA1, TS_5, TS_7, DI_10 e DE_10 são apresentados nas figuras 59, 60, 61, 62, e 63, respectivamente.

Há que se destacar a concentração de fragmentos cerâmicos com DI_10 e DI_10 em estacas no entorno do NSA3.

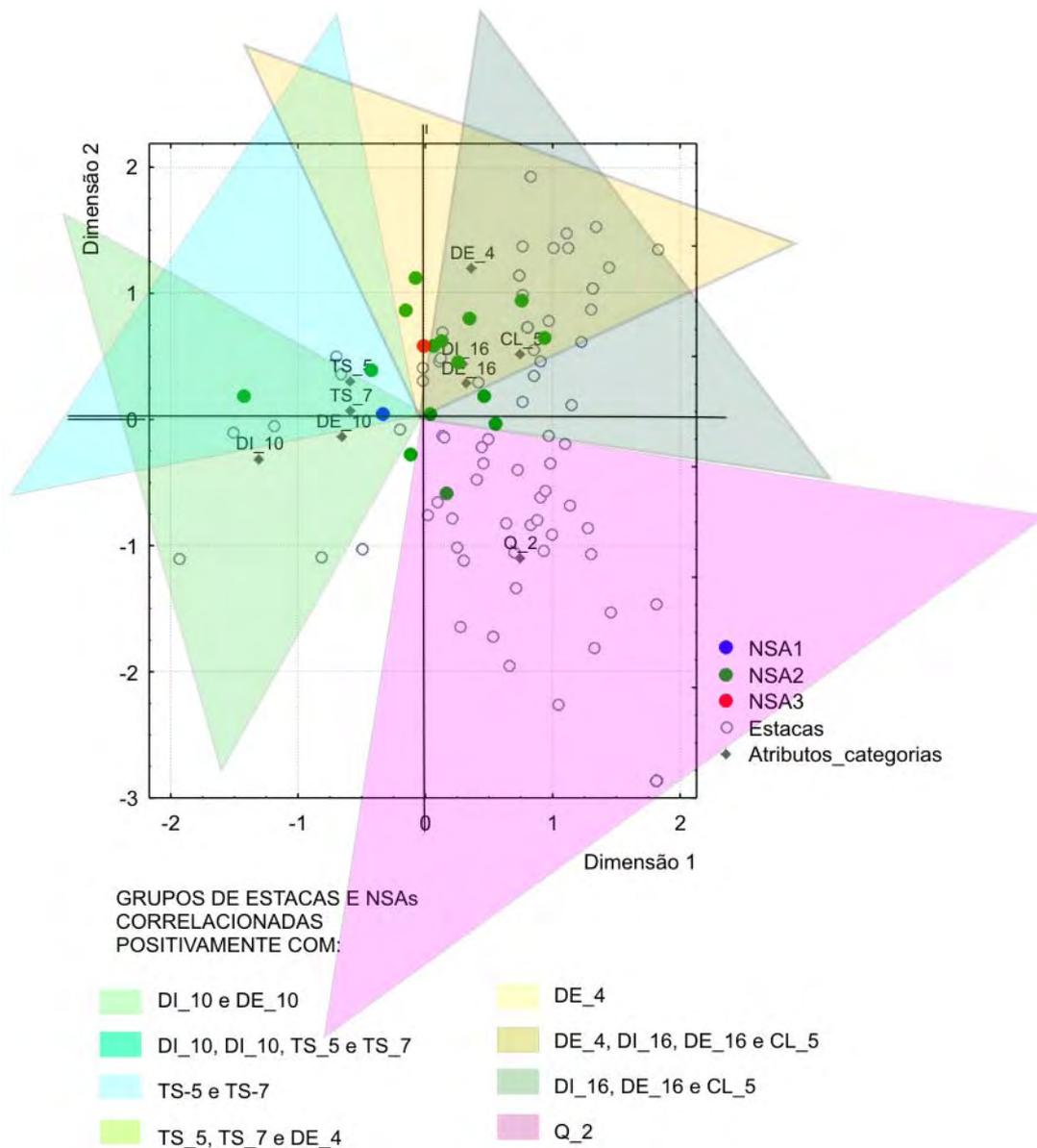


FIGURA 57 - Gráfico de dispersão das dimensões derivadas 1X2 das estacas e NSAs conjunto com atributos_categorias da tabela fraros, apresentando a correlação positiva entre estacas e NSAs e atributos_categorias.

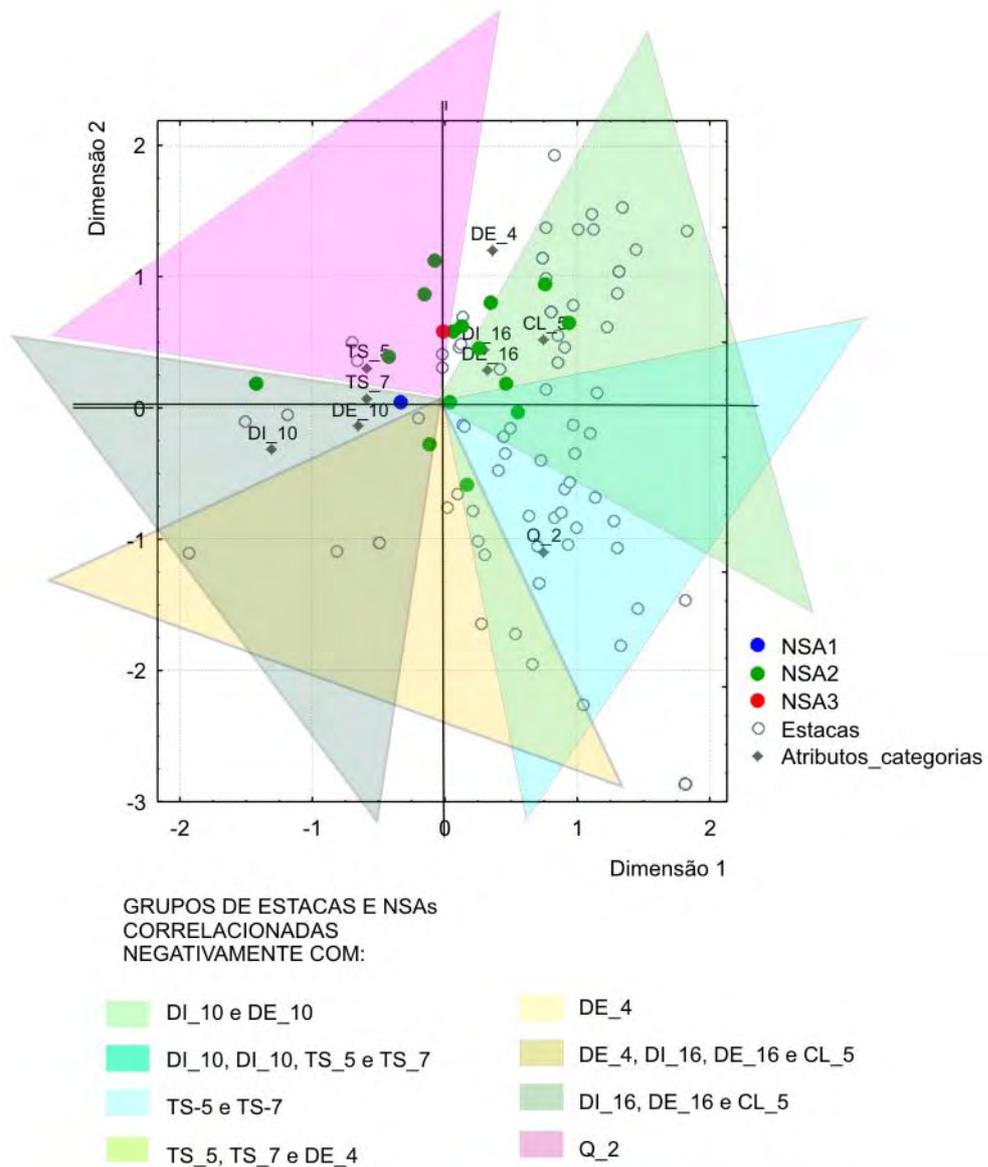


FIGURA 58 - Gráfico de dispersão das dimensões derivadas 1X2 das estacas e NSAs conjunto com atributos_categorias da tabela fraros, apresentando a correlação negativa entre estacas e NSAs e atributos_categorias.

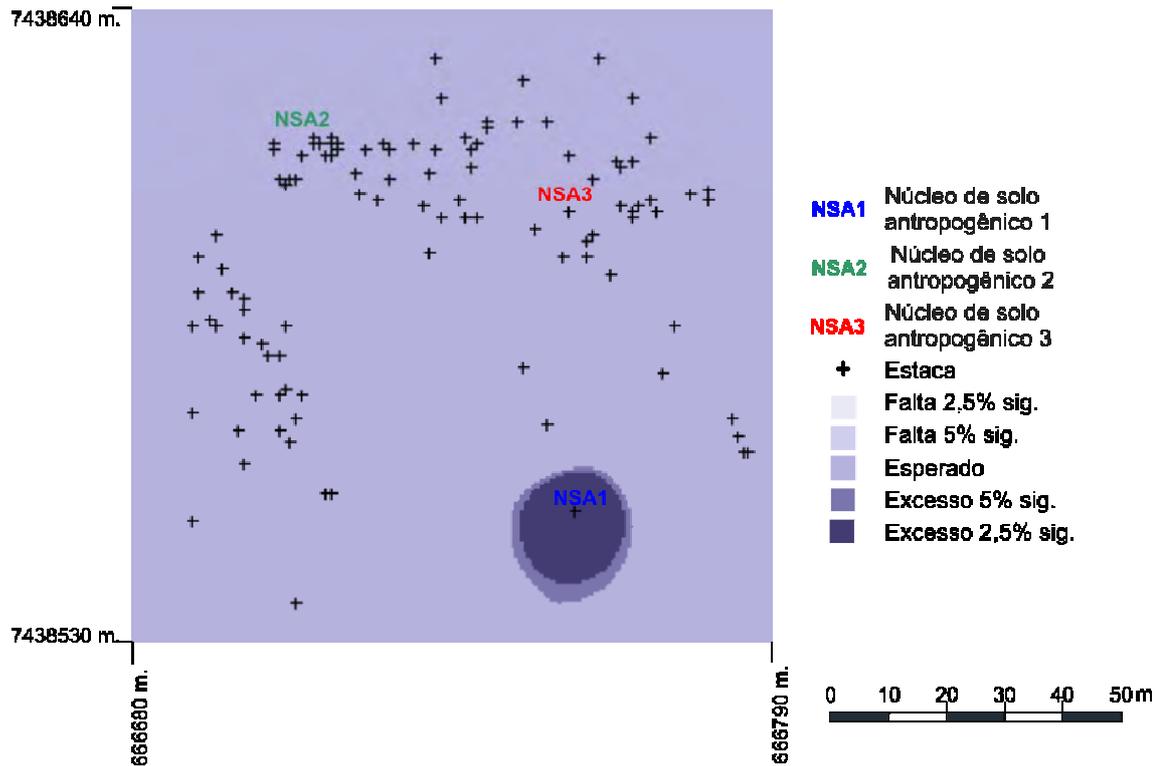


FIGURA 59 - Mapa temático do atributo tratamento superficial categoria 5 (TS_5).

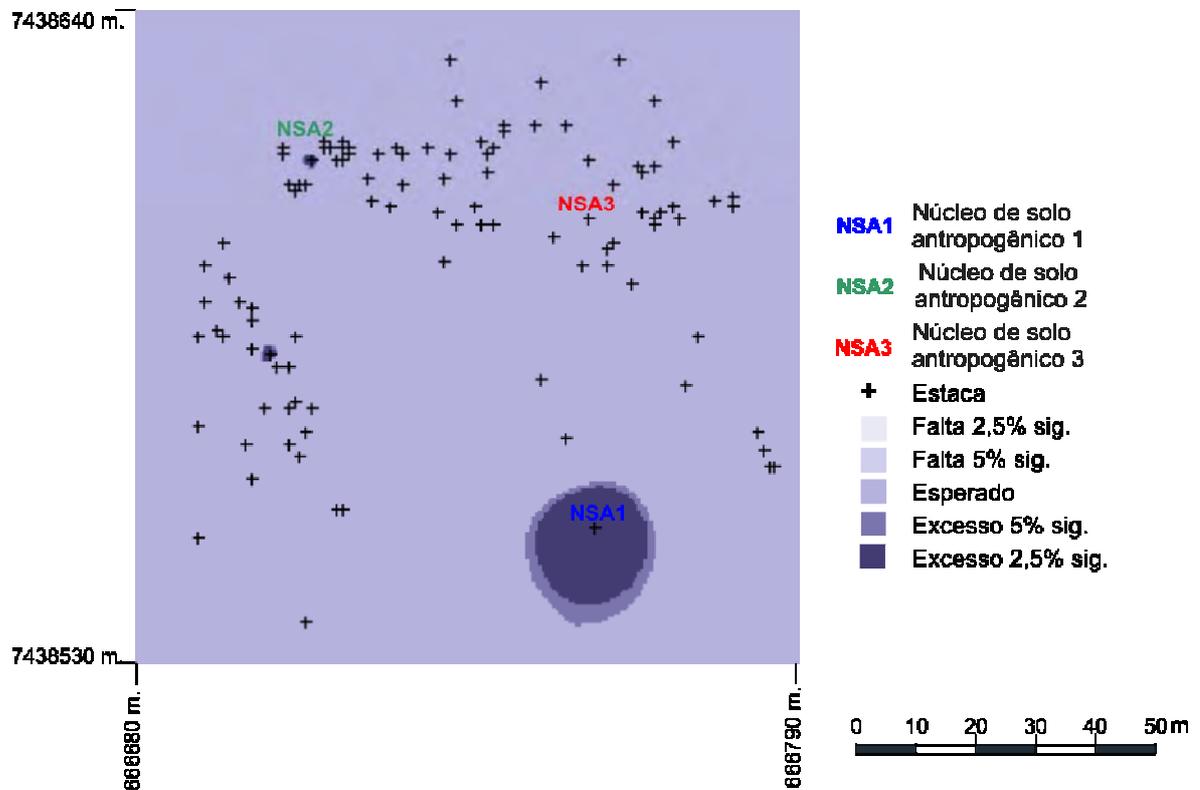


FIGURA 60 - Mapa temático do atributo tratamento superficial categoria 7 (TS_7).

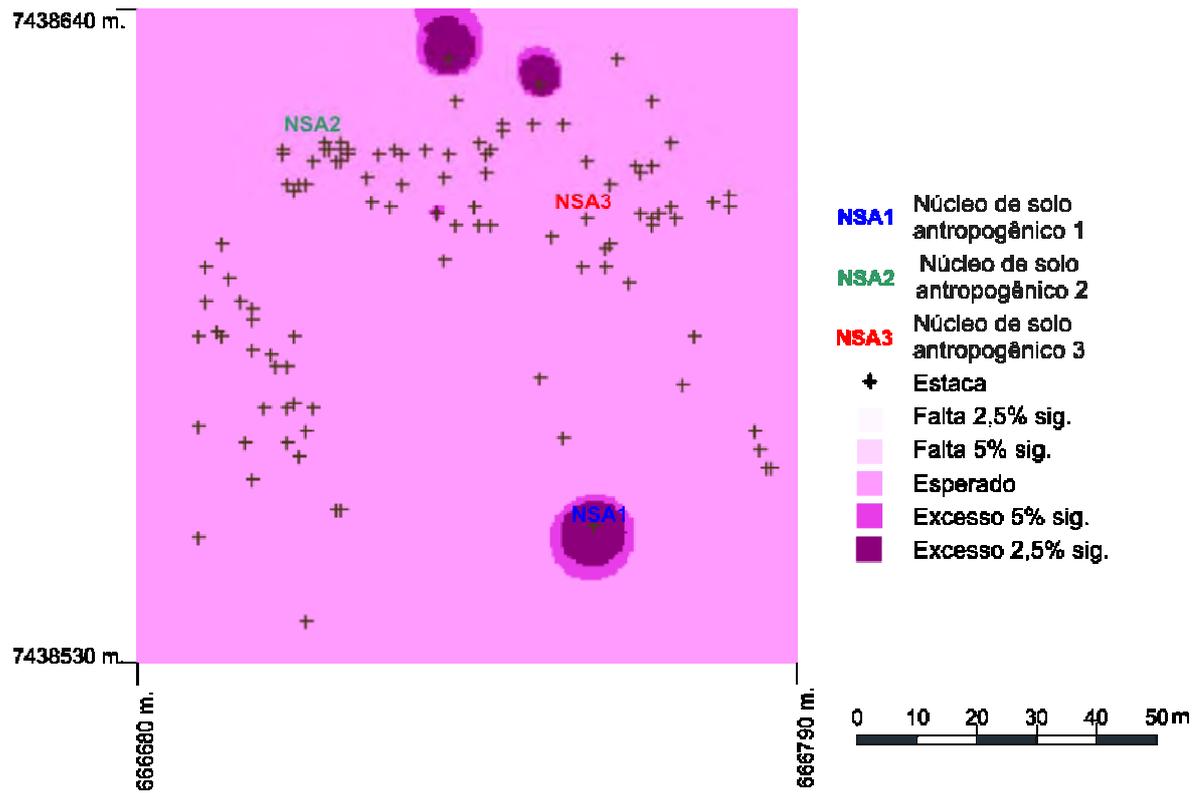


FIGURA 61 - Mapa temático do atributo decoração externa categoria 10 (DE_10).

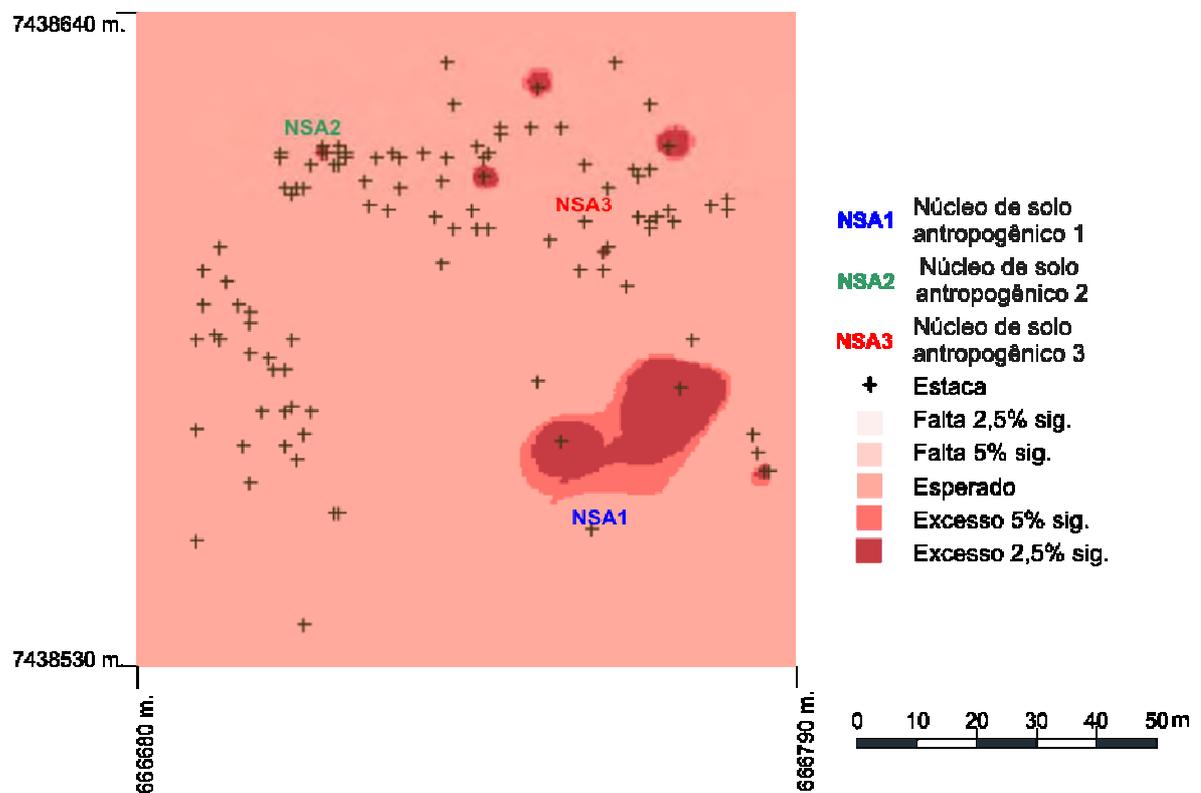


FIGURA 62 - Mapa temático do atributo decoração interna categoria 10 (DI_10).

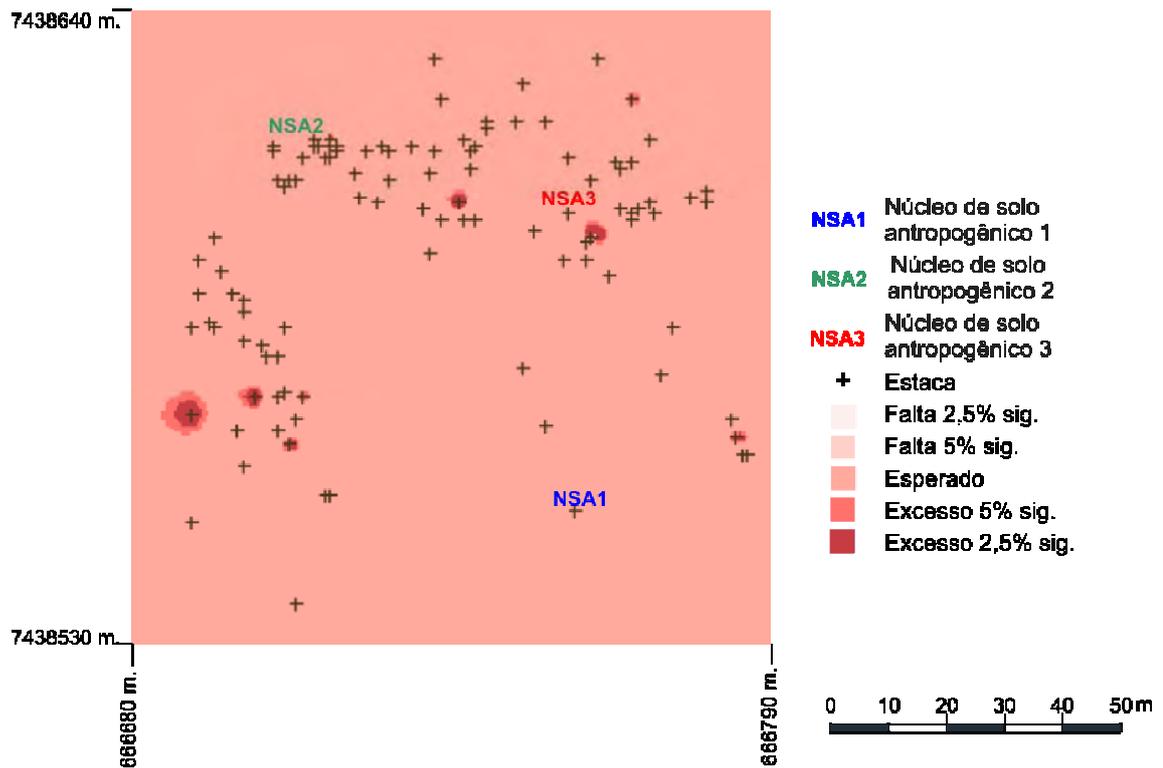


FIGURA 63 - Mapa temático do atributo decoração interna categoria 16 (DI_16).

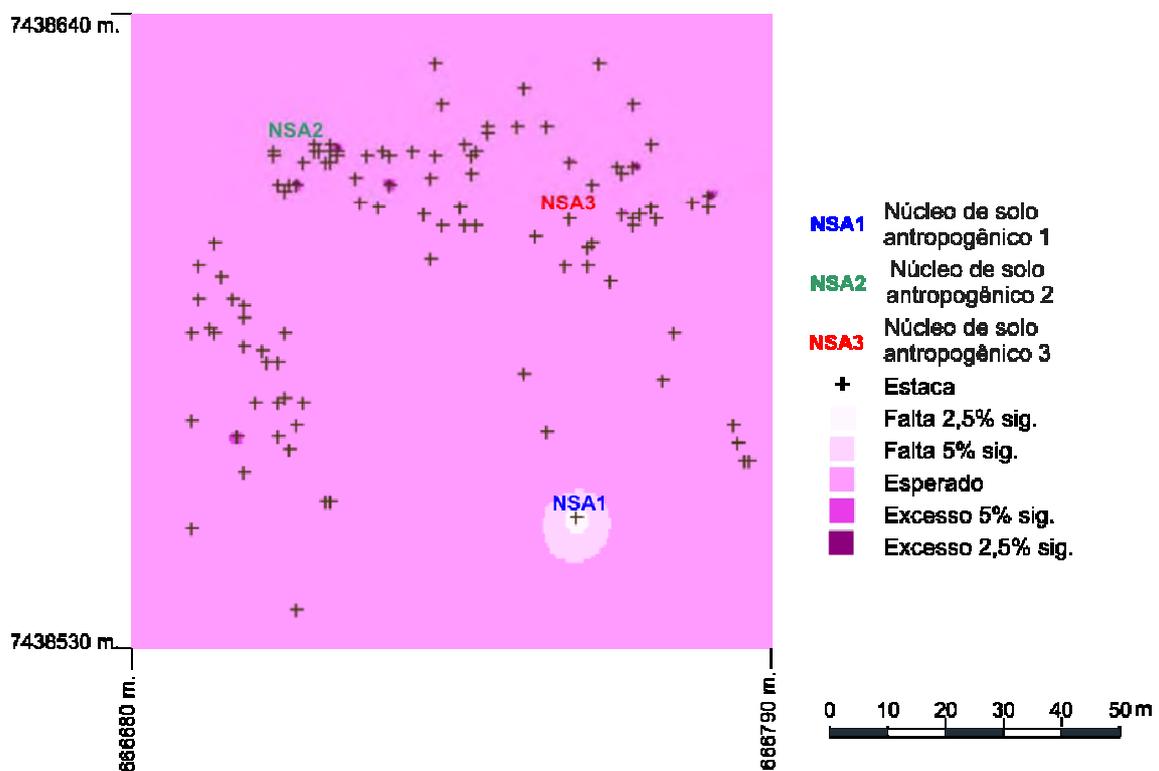


FIGURA 64 - Mapa temático do atributo decoração externa categoria 16 (DE_16).

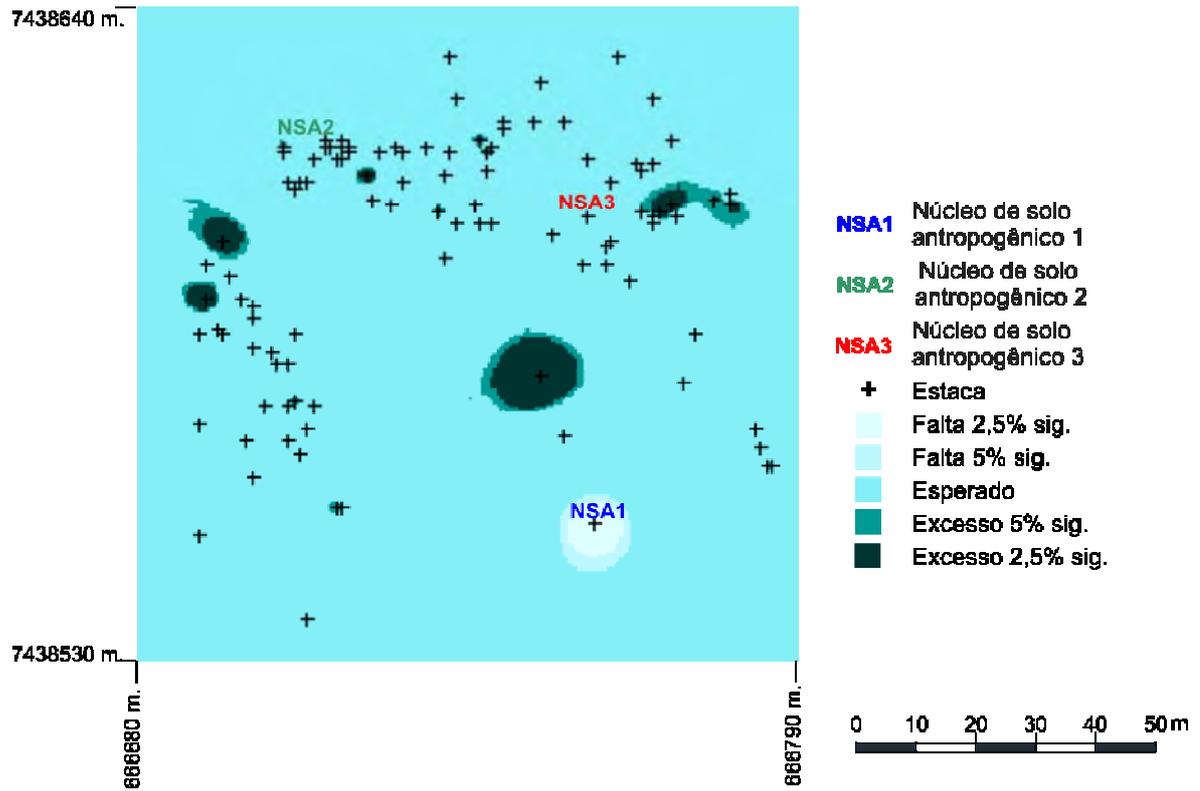


FIGURA 65 - Mapa temático do atributo classe categoria 5 (CL_5).

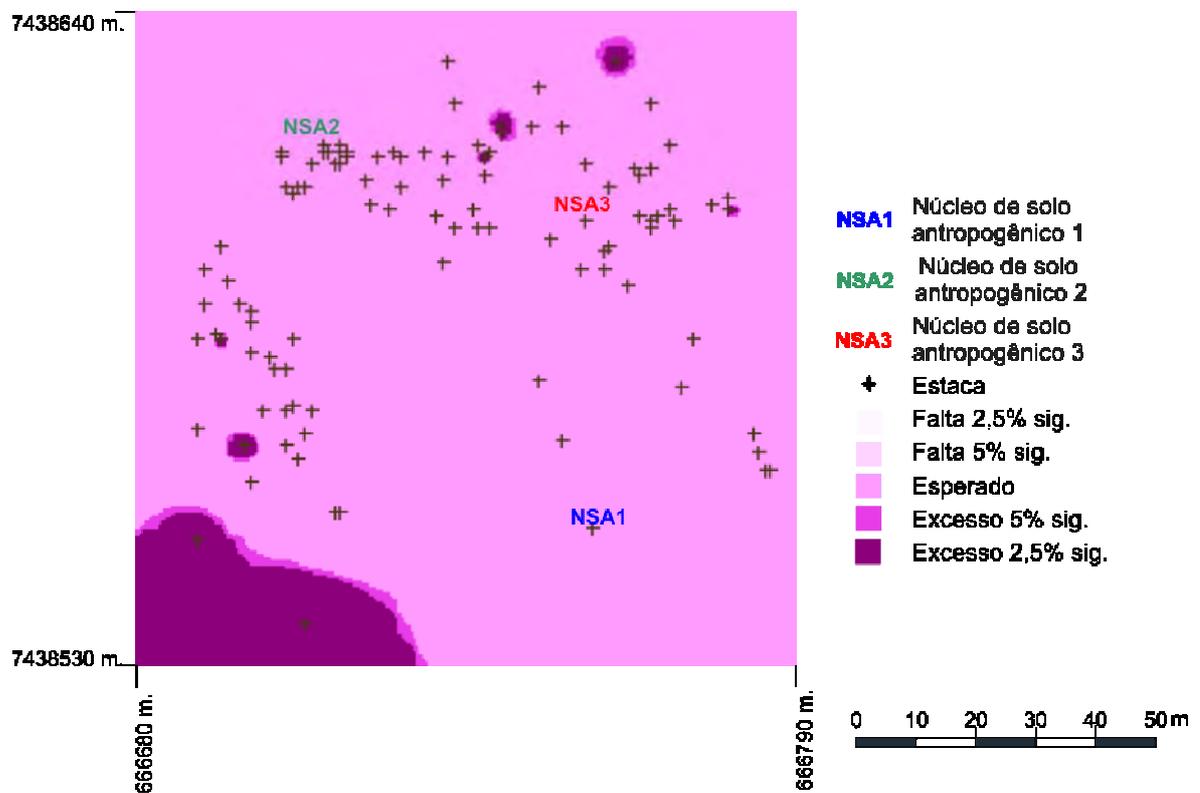


FIGURA 66 - Mapa temático do atributo decoração externa categoria 4 (DE_4).

Os atributos_categorias DI_16, DE_16, CL_5 encontram-se correlacionados ao

mesmo conjunto de estacas, parte do qual se correlaciona positivamente com DE_4. Os mapas temáticos dos resíduos padronizados desses atributos_categorias ilustram as figuras 63, 64, 65 e 66, respectivamente, nos quais observa-se também a correlação negativa que se estabelece entre o NSA1 e os atributos_categorias DE_16, e CL_5.

O mapa temático dos resíduos padronizados do atributo_categoria Q_2 que se correlaciona positivamente ao conjunto de estacas que não estabelece correlação positiva com nenhum outro atributo_categoria e correlação negativa com o NSA1, é apresentado na figura 67.

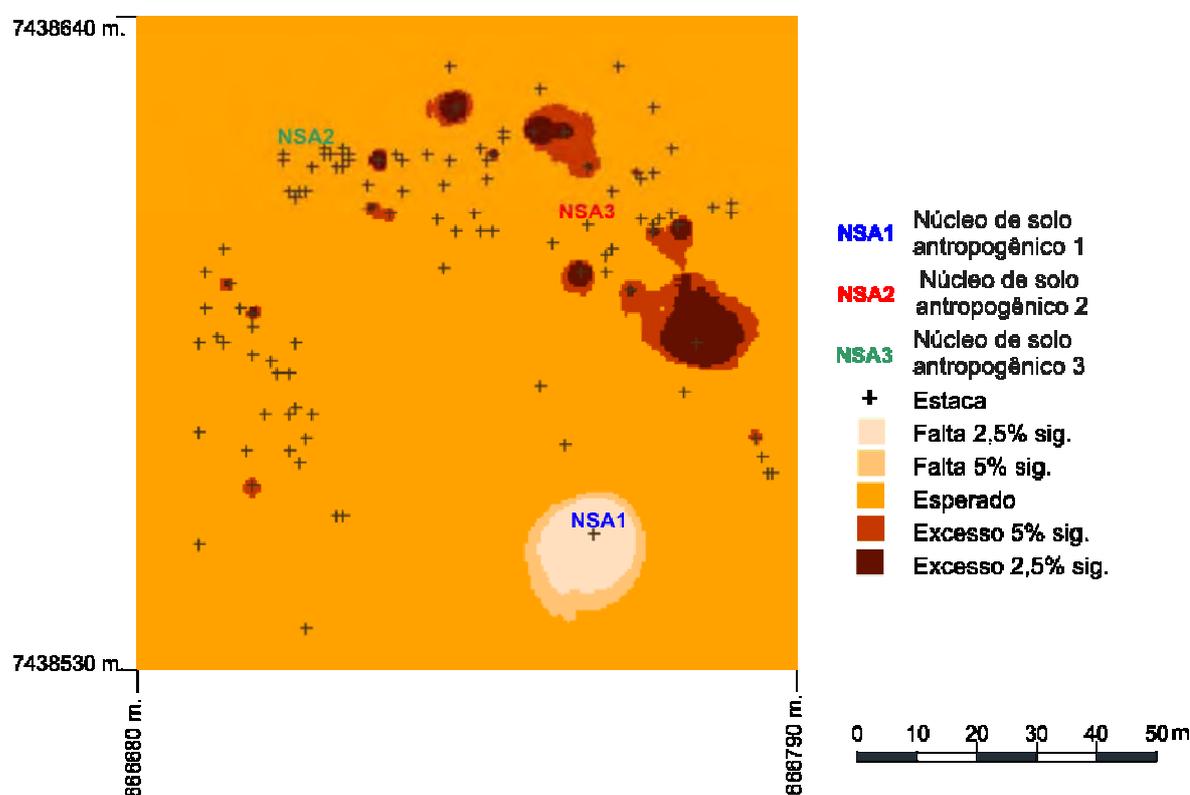


FIGURA 67 - Mapa temático do atributo queima categoria 2 (Q_2).

O estudo dos mapas temáticos dos resíduos padronizados dos atributos_categorias da tabela fraros dá destaque à área no entorno do NSA3 e ao NSA1.

O destaque ao entorno do NSA3 é dado pela diversidade de categorias de decoração interna e externa ali concentradas e ao NSA1 pela concentração de tratamentos superficiais diferenciados.

1.13.1.3 Tabela de contingência raríssimos

O exame do gráfico de linhas do perfil das linhas e do perfil marginal das linhas apresentado no apêndice A (mídia digital) no arquivo perfis – planilha perfis linha raríssimos indica não haver similaridade entre as estacas e NSAs quanto a distribuição dos atributos_categorias, existindo grande dispersão em relação ao perfil marginal das linhas, não sendo rara a ausência de determinados atributos_categorias nas estacas e NSAs.

Os perfis das colunas e o perfil marginal das colunas apresentados no gráfico de linhas que consta do apêndice A (mídia digital) no arquivo perfis – planilha perfis coluna raríssimos apresenta grande número de estacas onde se constata a ocorrência de um único atributo_categoria considerado raríssimo, como consequência se observa um perfil marginal das colunas próximo de zero, exceção feita aos NSA1 e NSA2.

A inércia total da tabela raríssimos é igual a 2,2077 significando grande dependência entre estacas e NSAs e atributos_categorias, ou seja, grande heterogeneidade na forma como esses atributos_categorias se apresentam nas estacas e NSAs.

O valor do teste do χ^2 é igual a 1744 sendo significativo a 2,5% de significância (o valor crítico da distribuição χ^2 com 1050 graus de liberdade é 1197, $p = 0,00$), isto é, as associações/relações entre as linhas e colunas, entre os elementos das linhas e entre os elementos das colunas detectadas pelo teste do χ^2 podem ser estudadas pelos resultados obtidos da aplicação da técnica da análise de correspondência com normalização canônica.

A tabela 3 apresenta os valores singulares, os autovalores, a inércia explicada a cada dimensão e a inércia explicada acumulada. As três primeiras dimensões contribuem com 49,3182% (em negrito) da inércia explicada total. A representação gráfica do comportamento das dimensões derivadasXautovalores é apresentada na figura 68.

Tabela 3 - Valores singulares, autovalores e inércia – tabela fraríssimos.

Dimensão	Valores Singulares	Autovalores	Inércia Explicada a cada dimensão	Inércia explicada acumulada
1	0,647293	0,418988	18,97862	18,9786
2	0,619400	0,383656	17,37821	36,3568
3	0,534927	0,286147	12,96141	49,3182
4	0,477261	0,227778	10,31750	59,6357
5	0,457618	0,209415	9,48571	69,1215
6	0,400006	0,160005	7,24763	76,3691
7	0,373506	0,139507	6,31915	82,6882
8	0,367236	0,134863	6,10877	88,7970
9	0,312087	0,097398	4,41178	93,2088
10	0,291775	0,085133	3,85620	97,0650
11	0,194585	0,037863	1,71506	98,7801
12	0,115503	0,013341	0,60429	99,3843
13	0,093274	0,008700	0,39408	99,7784
14	0,069940	0,004892	0,22157	100,0000
15	0,000000	0,000000	0,00000	100,0000

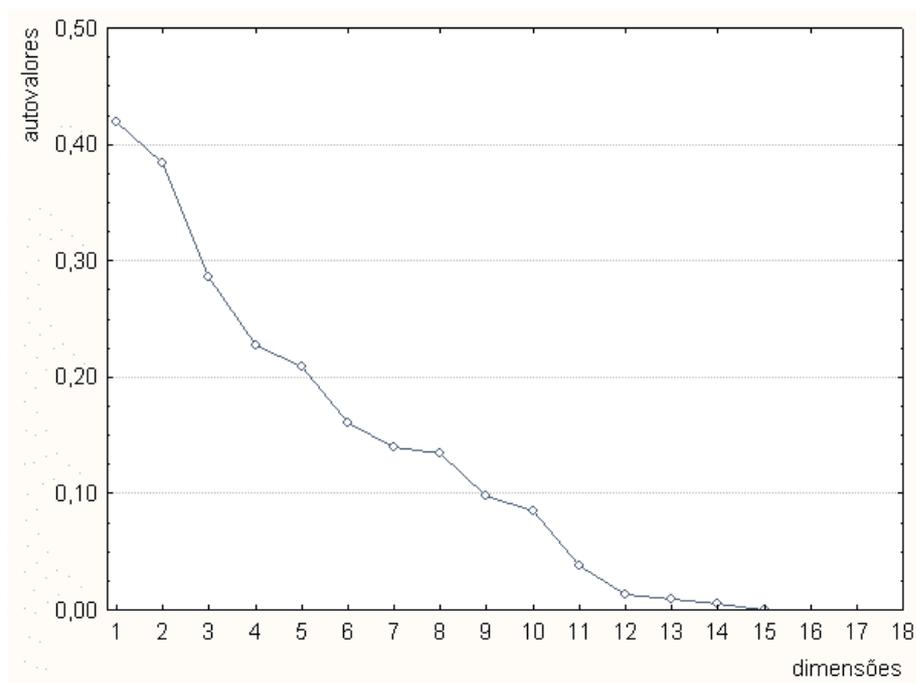


FIGURA 68 - Gráfico das dimensões derivadas – tabela fraríssimos

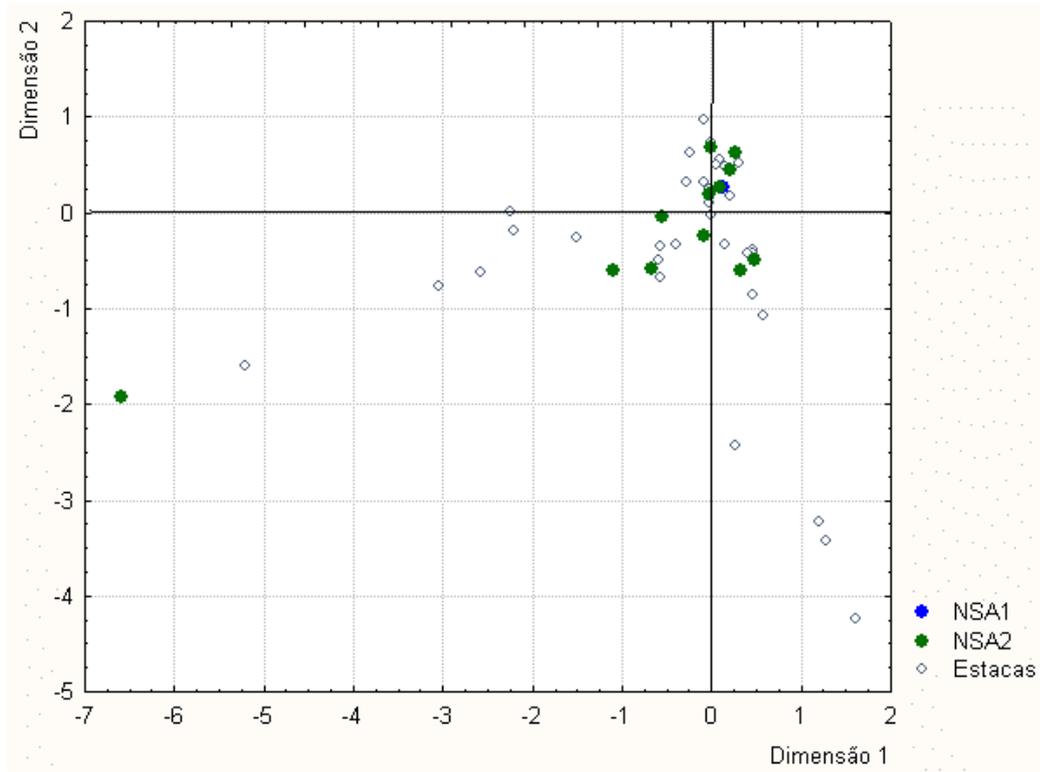


FIGURA 69 - Gráfico de dispersão das dimensões derivadas 1X2 das estacas e NSAs da tabela fraríssimos

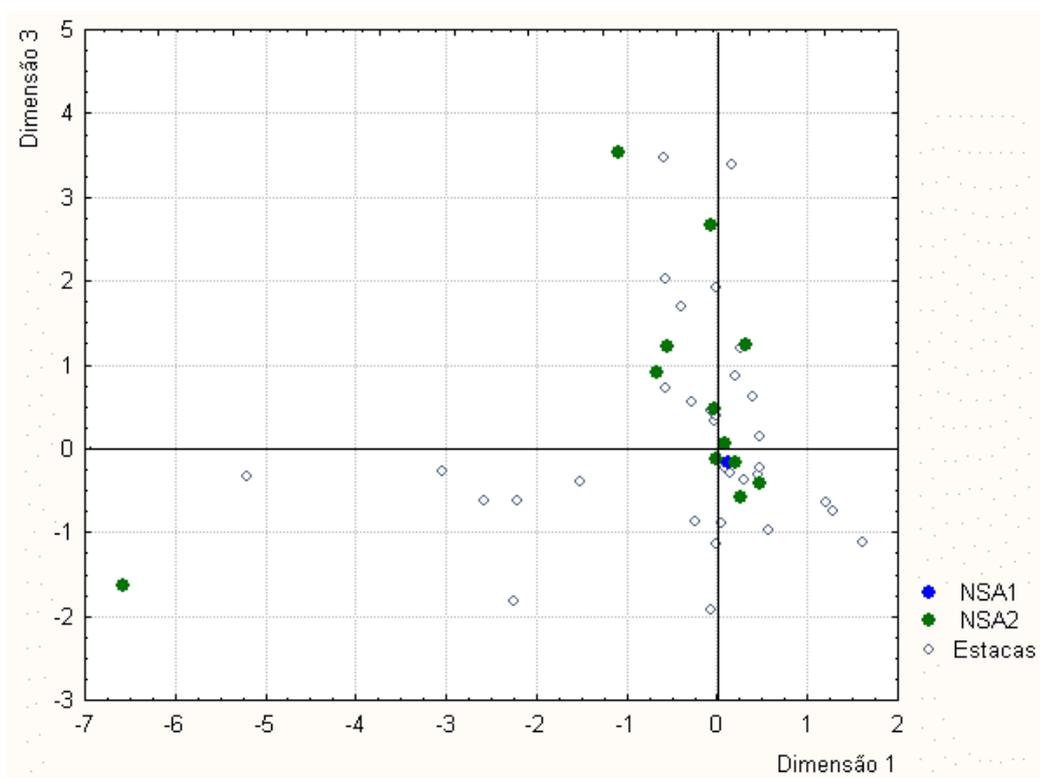


FIGURA 70 - Gráfico de dispersão das dimensões derivadas 1X3 das estacas e NSAs da tabela fraríssimos

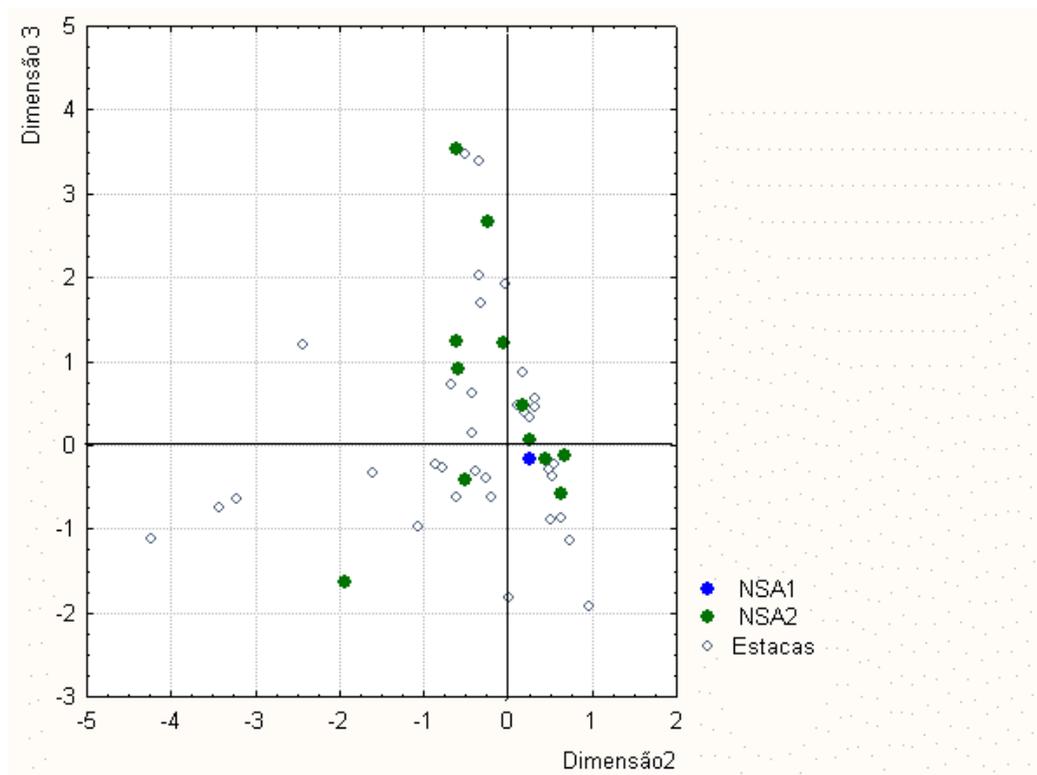


FIGURA 71 - Gráfico de dispersão das dimensões derivadas 2X3 das estacas e NSAs da tabela fraríssimos

O estudo das dimensões derivadas 1, 2 e 3 das estacas e NSAs, ilustrado na figura 69, dimensão 1X2, 70, dimensão 1X3, e 71, dimensão 2X3, mostra a dispersão das estacas e NSAs ao longo dos três eixos derivados definindo três grupos de estacas: um associado à dimensão derivada 1, outro à dimensão derivada 2 e outro à dimensão derivada 3. Consta-se a presença de estacas que se distanciam muito do centro de massa da distribuição (“*outliers*”).

Os gráficos das dimensões derivadas dos atributos_categorias em suas dimensões derivadas 1X2X3 que ilustra a figura 72, tornam clara a presença de atributos_categorias que se configuram como “outlier”, são eles: DE_2, DE_5 e TS_20. Os mapas temáticos dos resíduos padronizados desses atributos_categorias que ilustram as figuras 73, 74 e 75, respectivamente, destacam a falta de fragmentos cerâmicos com DE_2 e DE_5 no NSA1, bem como pontos de concentração conjunta dos mesmos no entorno dos NSA2 e NSA3.

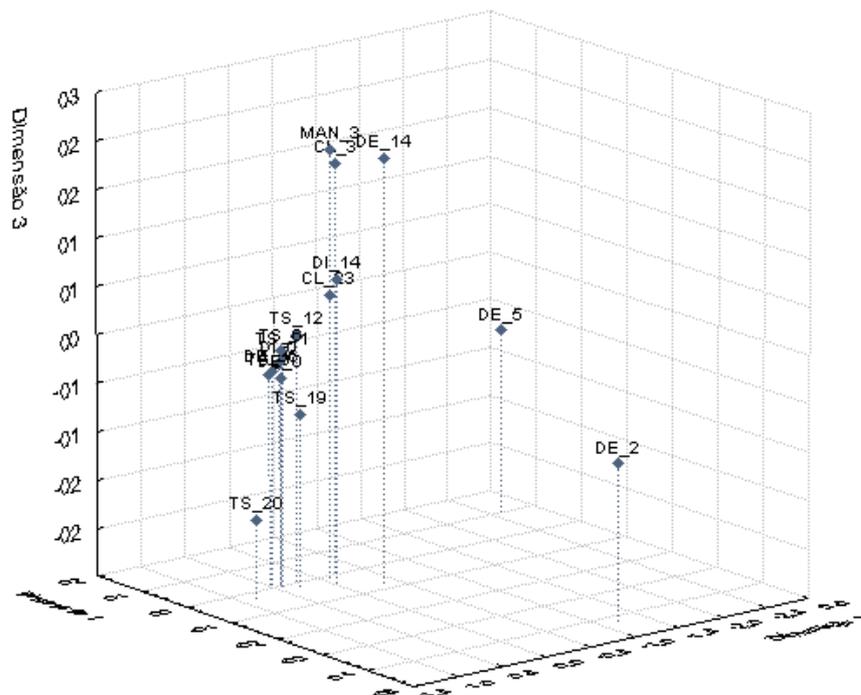


FIGURA 72 - Gráfico de dispersão das dimensões derivadas 1X2X3 dos atributos_categorias da tabela fraríssimos

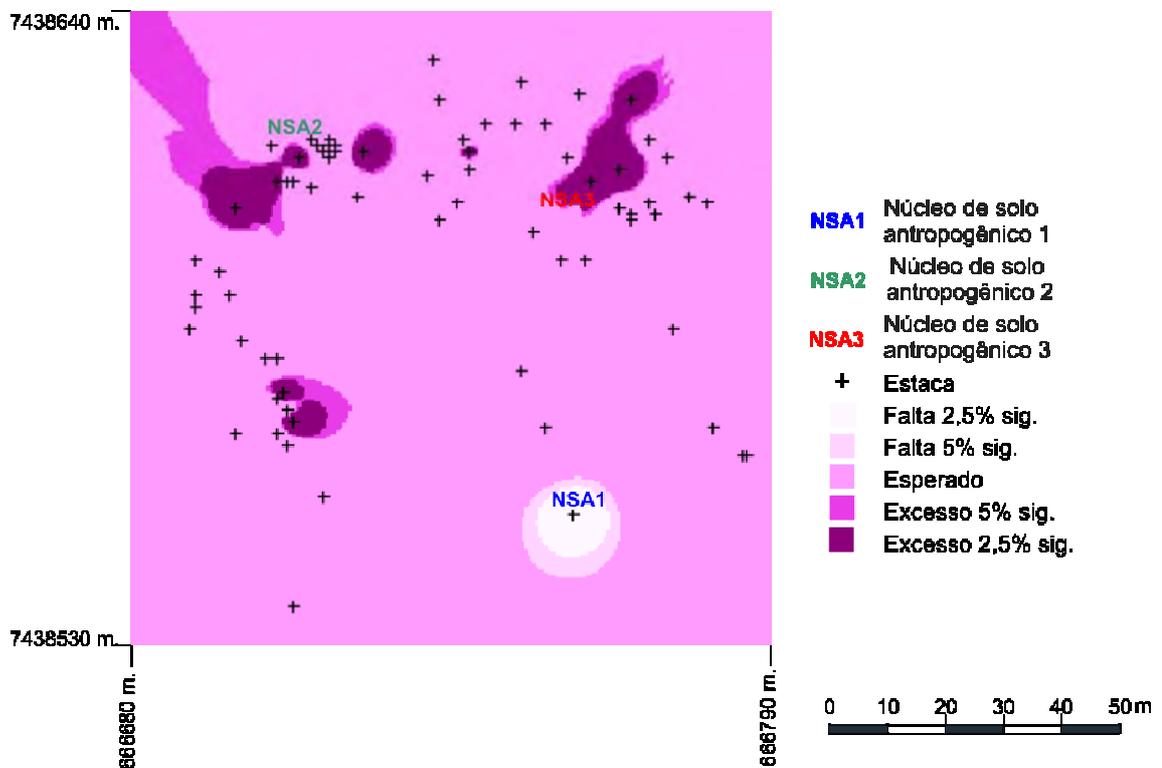


FIGURA 73 - Mapa temático do atributo decoração externa categoria 2 (DE_2).

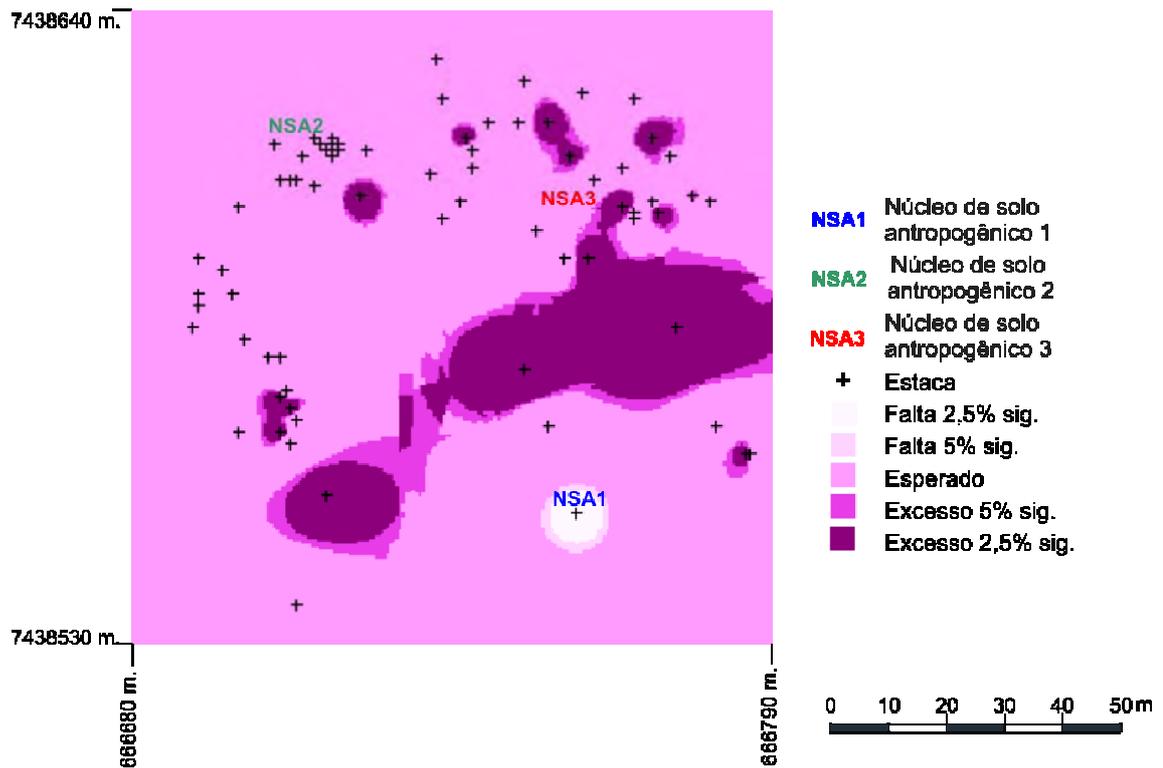


FIGURA 74 - Mapa temático do atributo decoração externa categoria 5 (DE_5).

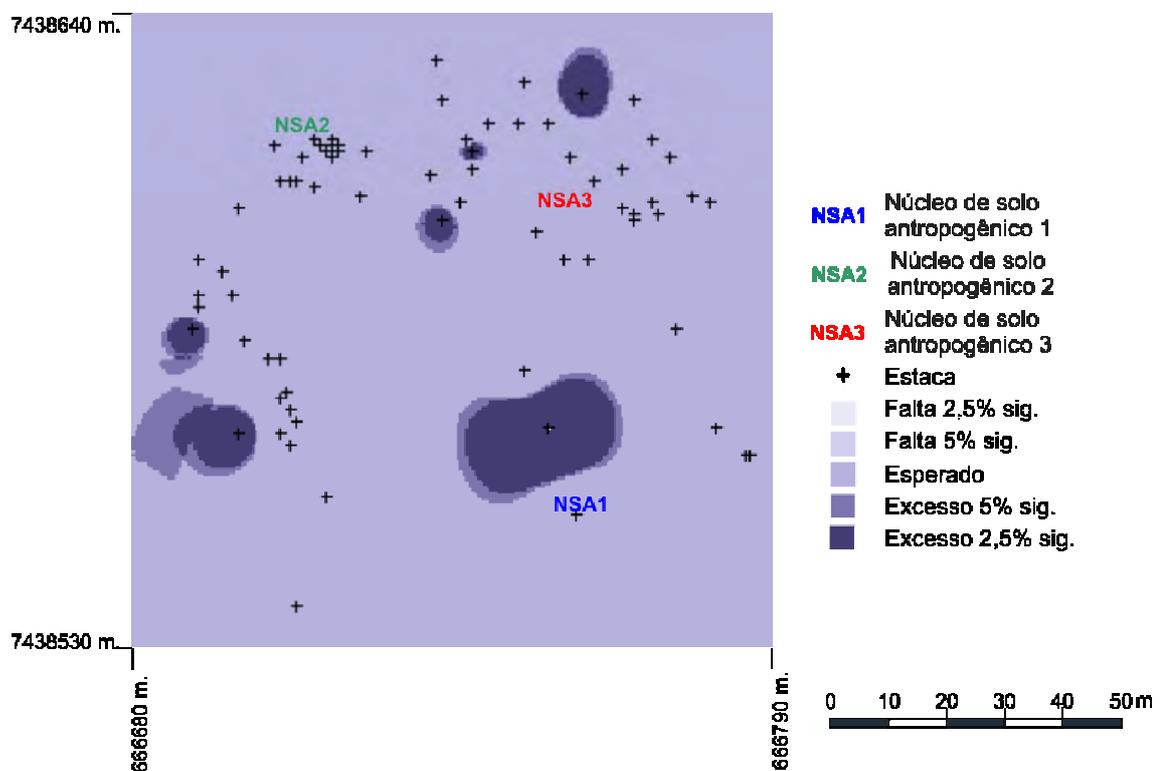


FIGURA 75 - Mapa temático do atributo tratamento superficial categoria 20 (TS_20).

Os atributos_categorias DI_0, DE_0, DE_36, TS_0 e TS_11 são os que mais se aproximam do centro de massa da distribuição. Os mapas temáticos dos resíduos padronizados desses atributos_categorias são apresentados nas figuras 76, 77, 78, 79 e 80, respectivamente, nos quais se evidencia a concentração conjunta no NSA1.

A análise dos demais atributos_categorias revela a associação de MAN_3, CL_3 e DE_14, isto é, fragmentos de base tendem a ser modelados à mão e ter decoração externa de engobo vermelho/branco; de CL_23 e DI_14, fragmentos de parede/borda angular tendem a ter decoração interna de engobo vermelho/branco. A análise das relações/associações desses atributos_categorias bem como de TS_6, TS_12 e TS_19 com as estacas e NSAs será efetuada no gráfico conjunto das estacas e NSAs com os atributos_categorias nas dimensões derivadas 1X2 apresentado nas figuras 81 e 82.

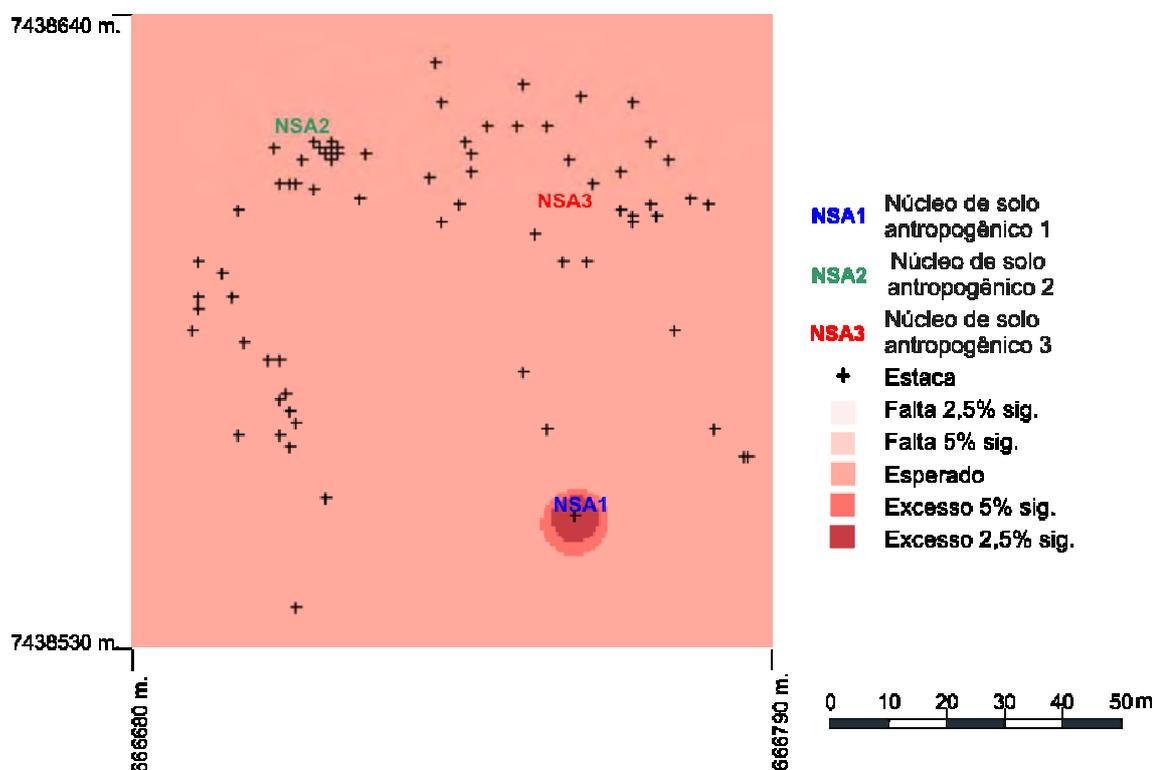


FIGURA 76 - Mapa temático do atributo decoração interna categoria 0 (DI_0).

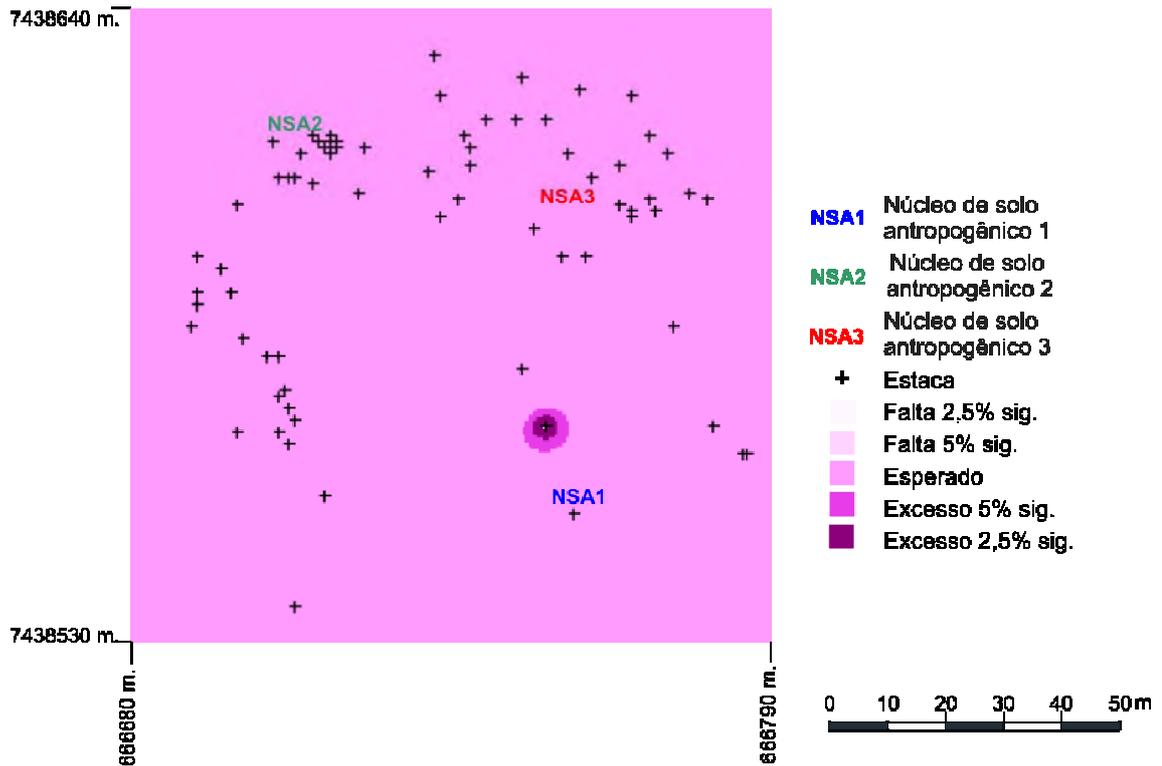


FIGURA 77 - Mapa temático do atributo decoração externa categoria 0 (DE_0).

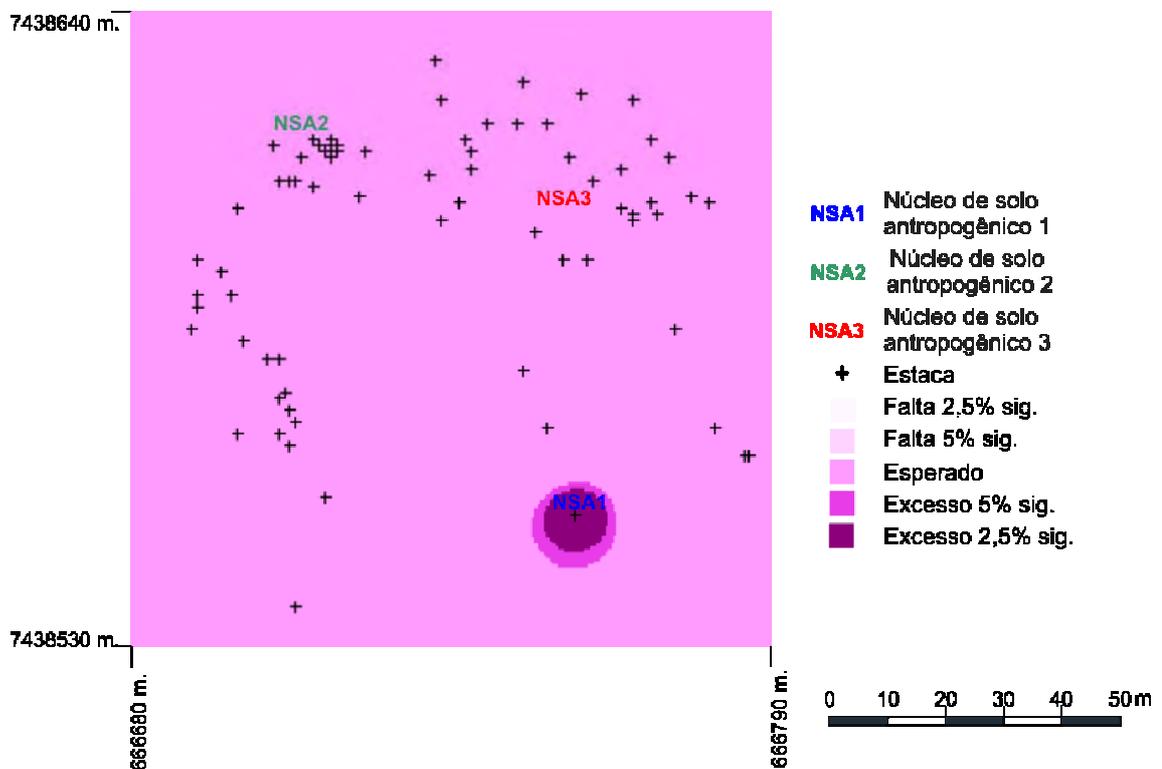


FIGURA 78 - Mapa temático do atributo decoração externa categoria 36 (DE_36).

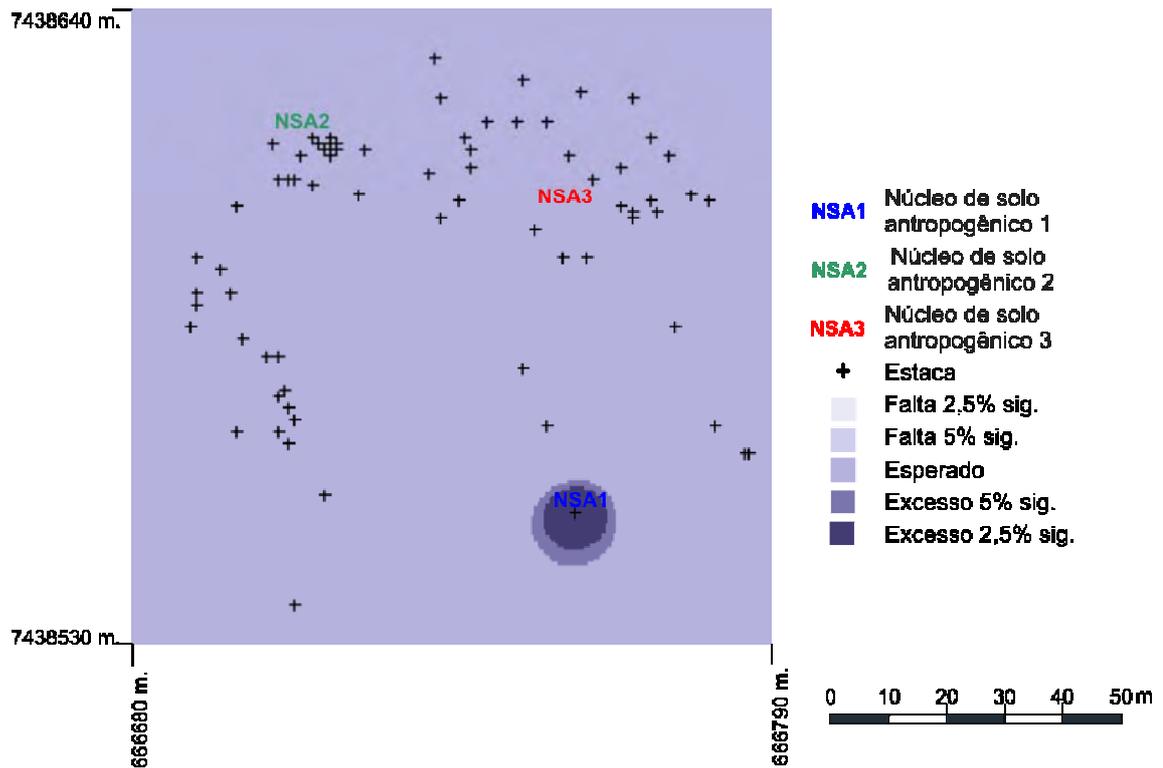


FIGURA 79 - Mapa temático do atributo tratamento superficial categoria 0 (TS_0).

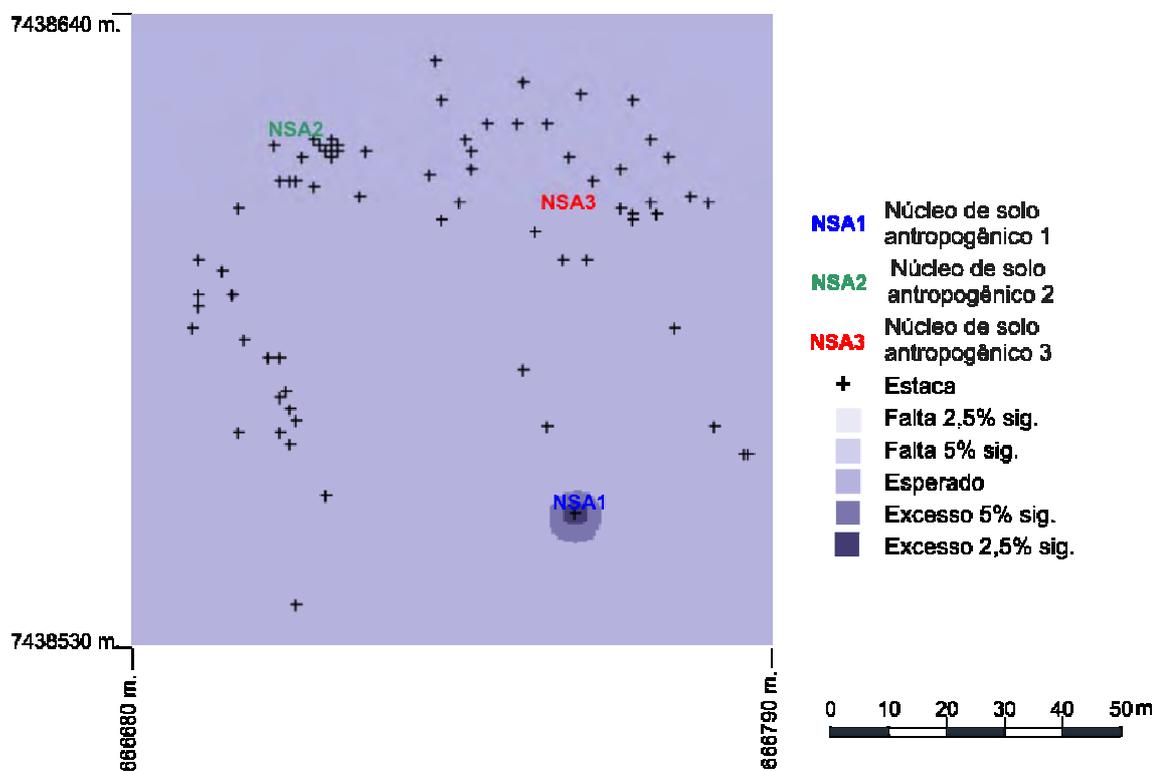


FIGURA 80 - Mapa temático do atributo tratamento superficial categoria 11 (TS_11).

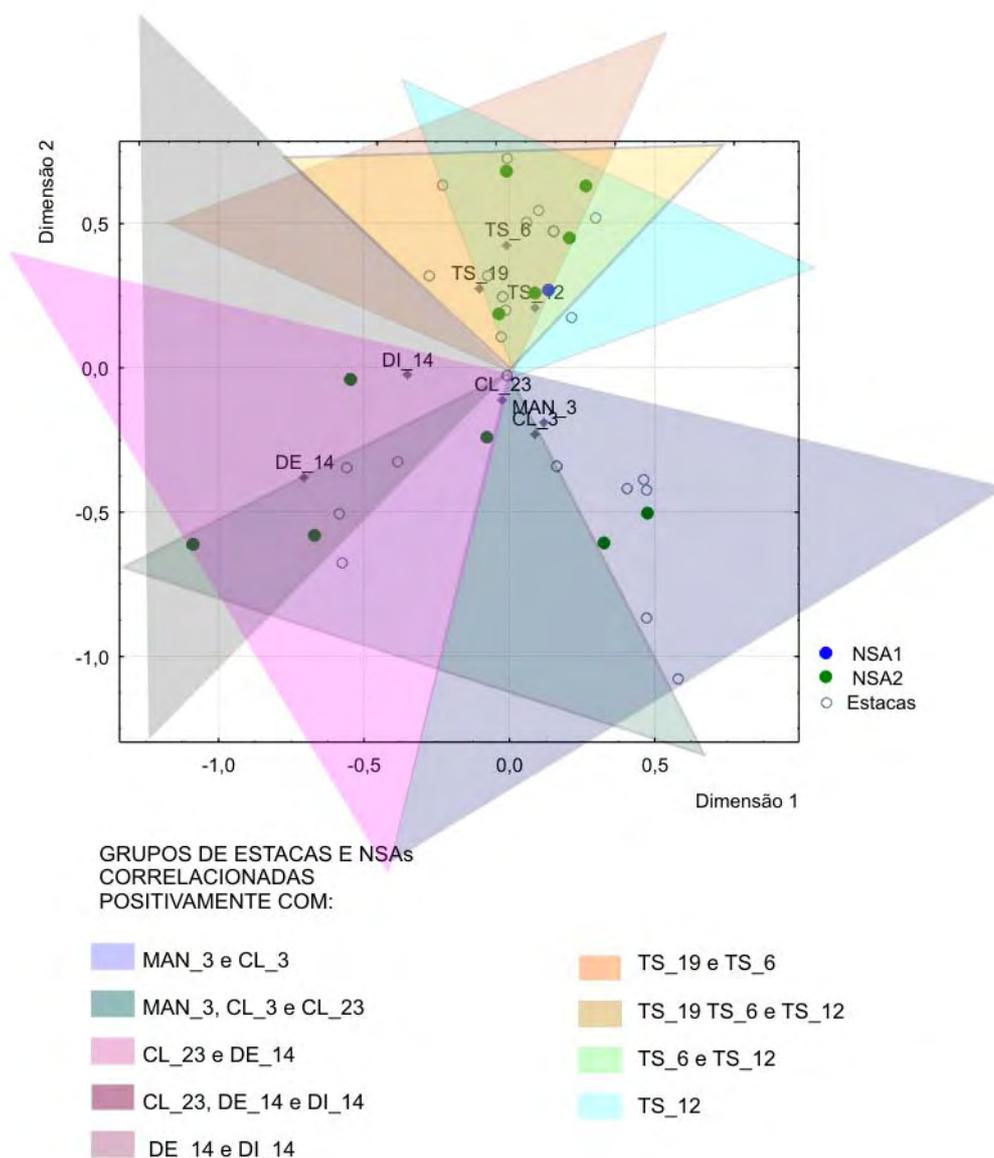


FIGURA 81 - Gráfico de dispersão das dimensões derivadas 1X2 das estacas e NSAs conjunto com atributos_categorias da tabela fraríssimos, apresentando a correlação positiva entre estacas e NSAs e atributos_categorias.

Das figuras 81 e 82 depreende-se que:

- do grupo de estacas que estabelece correlação positiva com MAN_3 e CL_3, parte se correlaciona positivamente com CL_23 e estabelece correlação negativa com TS_6, TS_9 e TS_12. Os mapas temáticos dos resíduos padronizados dos atributos_categorias MAN_3, CL_3 e CL_23 são apresentados nas figuras 83, 84 e 85, respectivamente, nos quais se observa a concentração desses

atributos_categorias que se desenvolve no NSA2 e na meia distância do NSA2 e NSA3, bem como a falta de MAN_3 e CL_3 no NSA1.

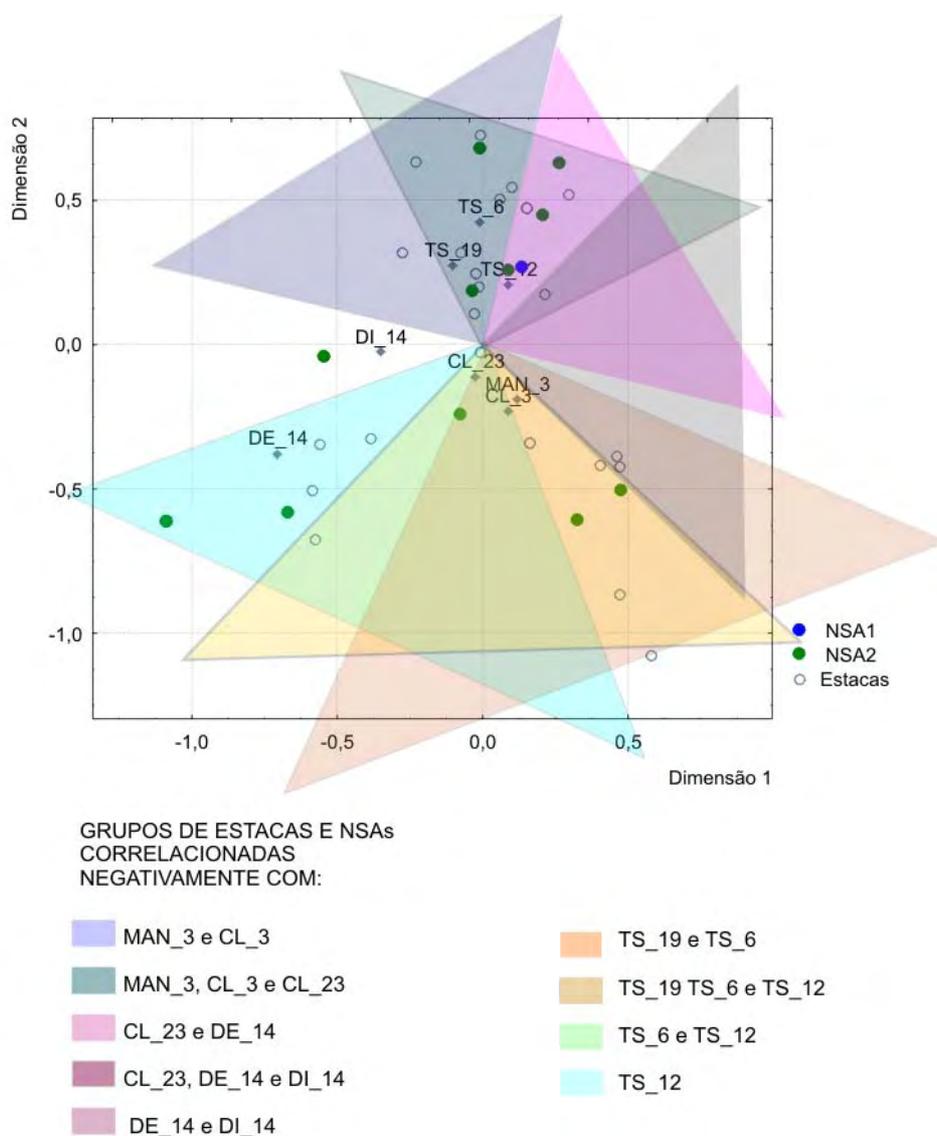


FIGURA 82 - Gráfico de dispersão das dimensões derivadas 1X2 das estacas e NSAs conjunto com atributos_categorias da tabela fraríssimos, apresentando a correlação negativa entre estacas e NSAs e atributos_categorias.

- DE_14 e DI_14 estão correlacionados positivamente ao mesmo grupo de estacas, parte das quais se correlaciona positivamente com CL_23, e negativamente com TS_12. O mapa temático dos resíduos padronizados de DE_14 e DI_14 ilustra as figuras 86 e 87, respectivamente, e apresenta pontos de concentração conjunta no NSA2.

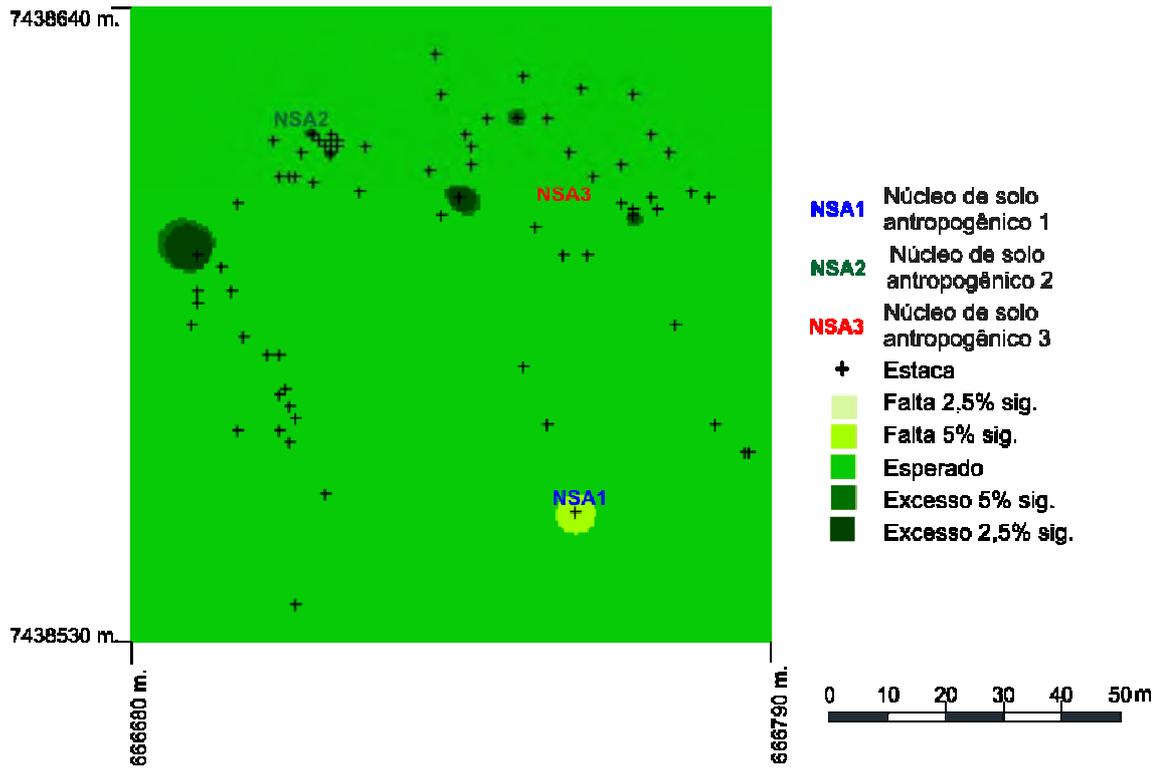


FIGURA 83 - Mapa temático do atributo técnica de manufatura categoria 3 (MAN_3).

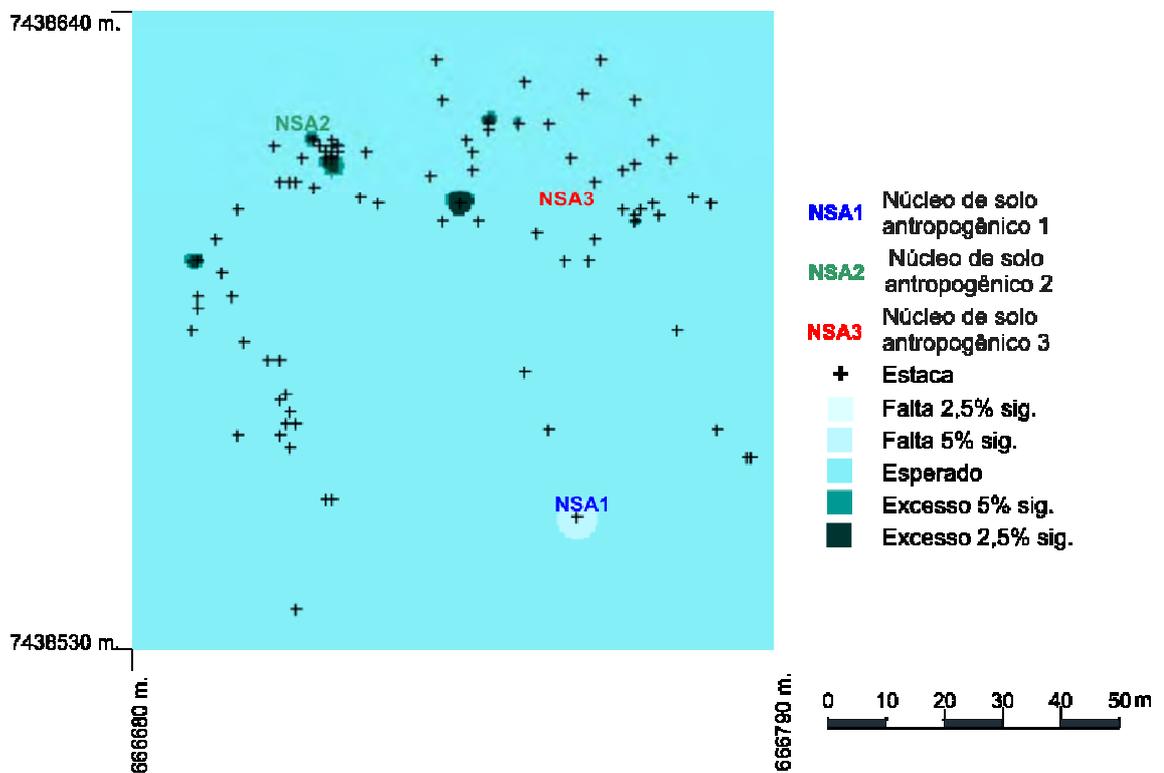


FIGURA 84 - Mapa temático do atributo classe categoria 3 (CL_3).

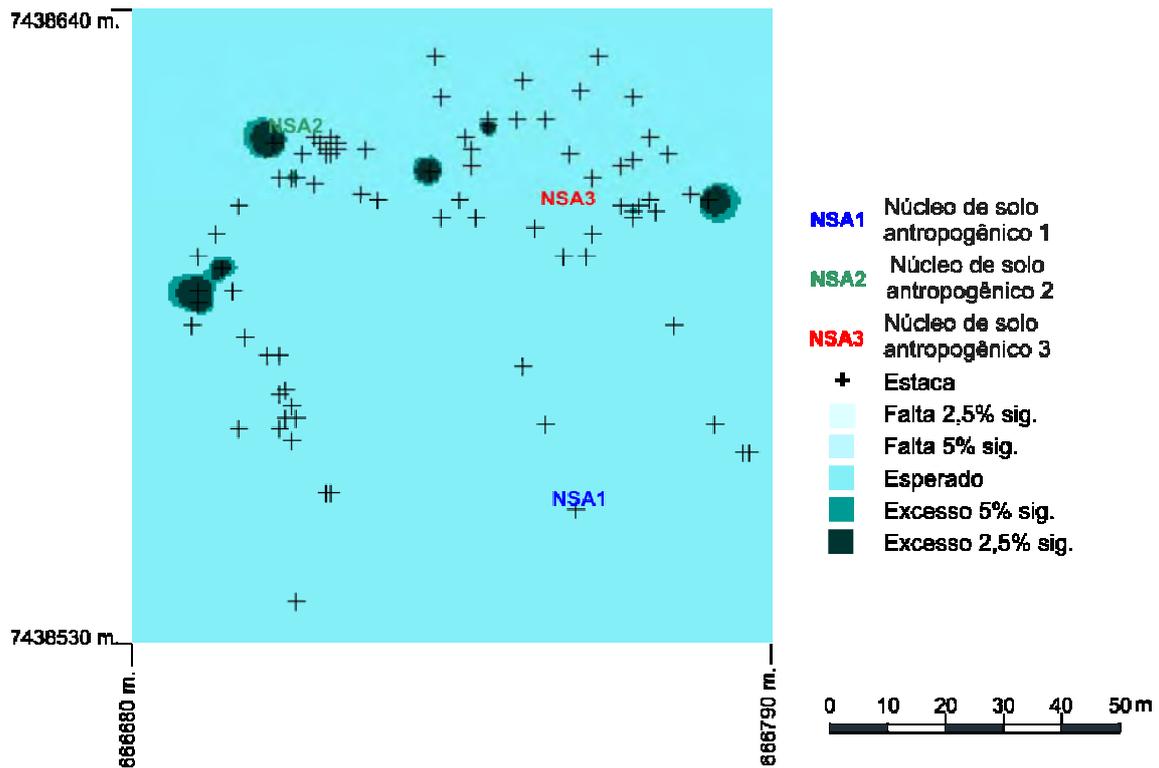


FIGURA 85 - Mapa temático do atributo classe categoria 23 (CL_23).

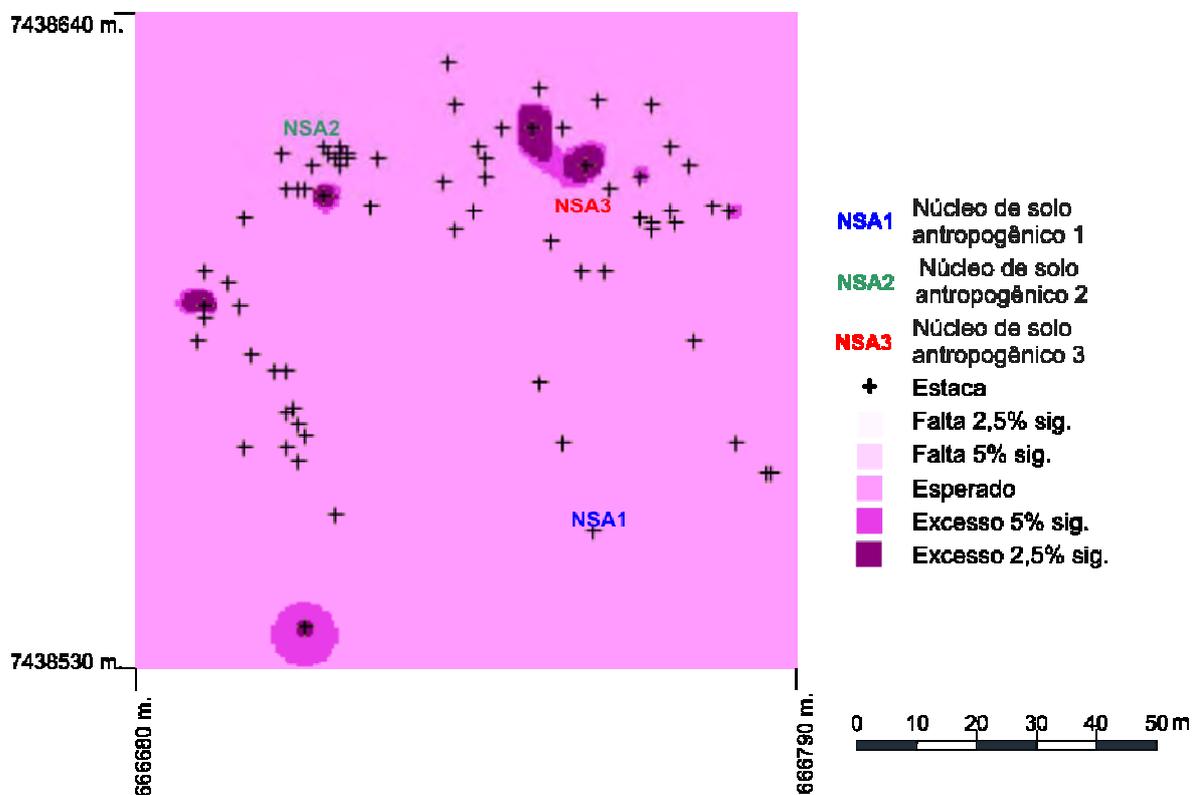


FIGURA 86 - Mapa temático do atributo decoração externa categoria 14 (DE_14).

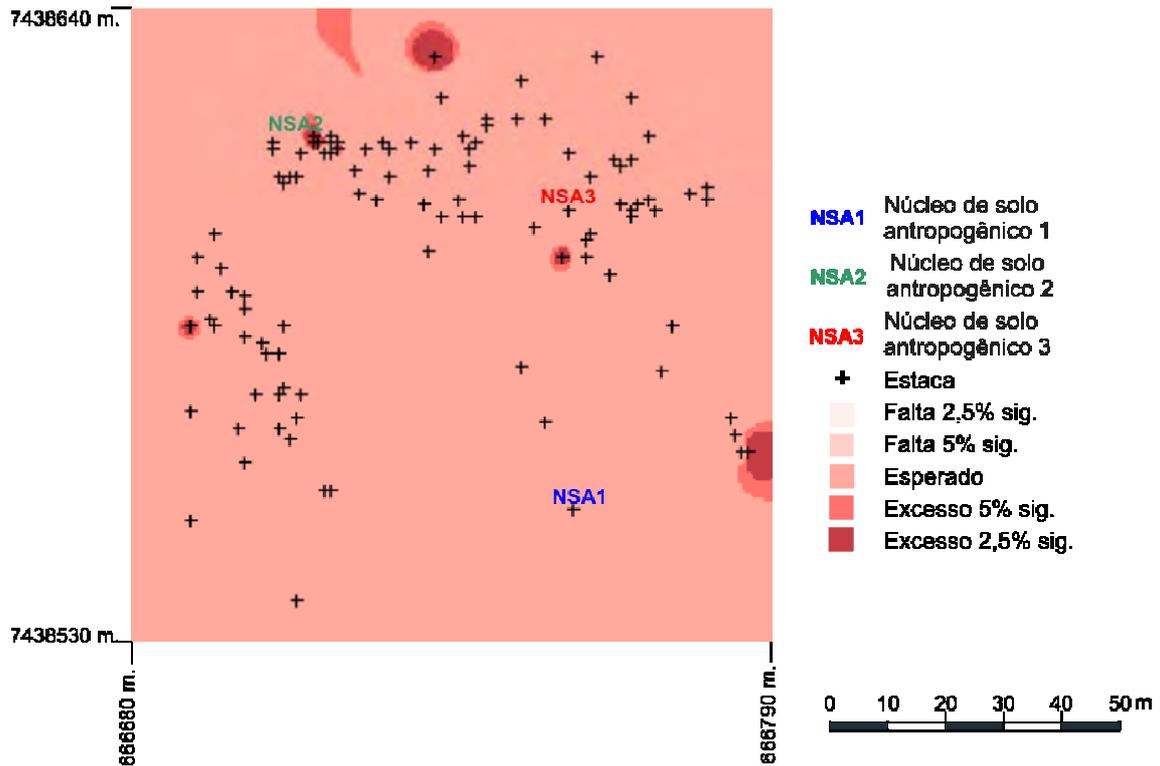


FIGURA 87 - Mapa temático do atributo decoração interna categoria 14 (DI_14).

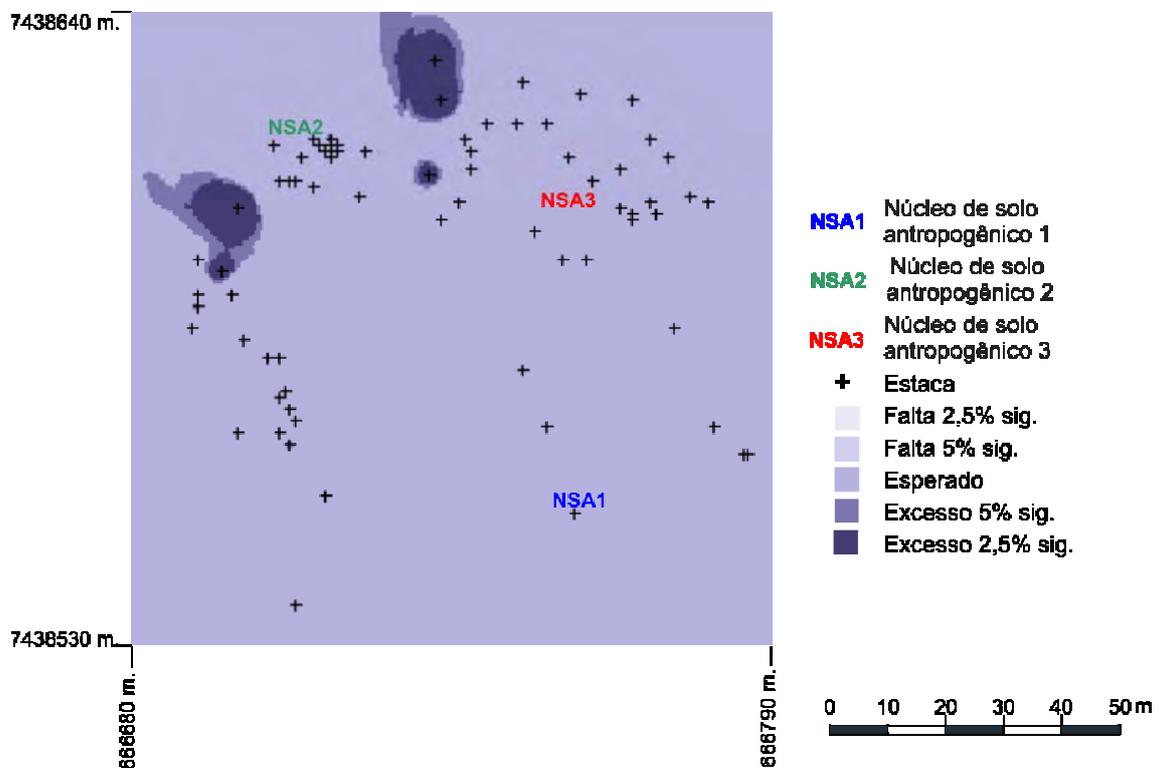


FIGURA 88 - Mapa temático do atributo tratamento superficial categoria 6 (TS_6).

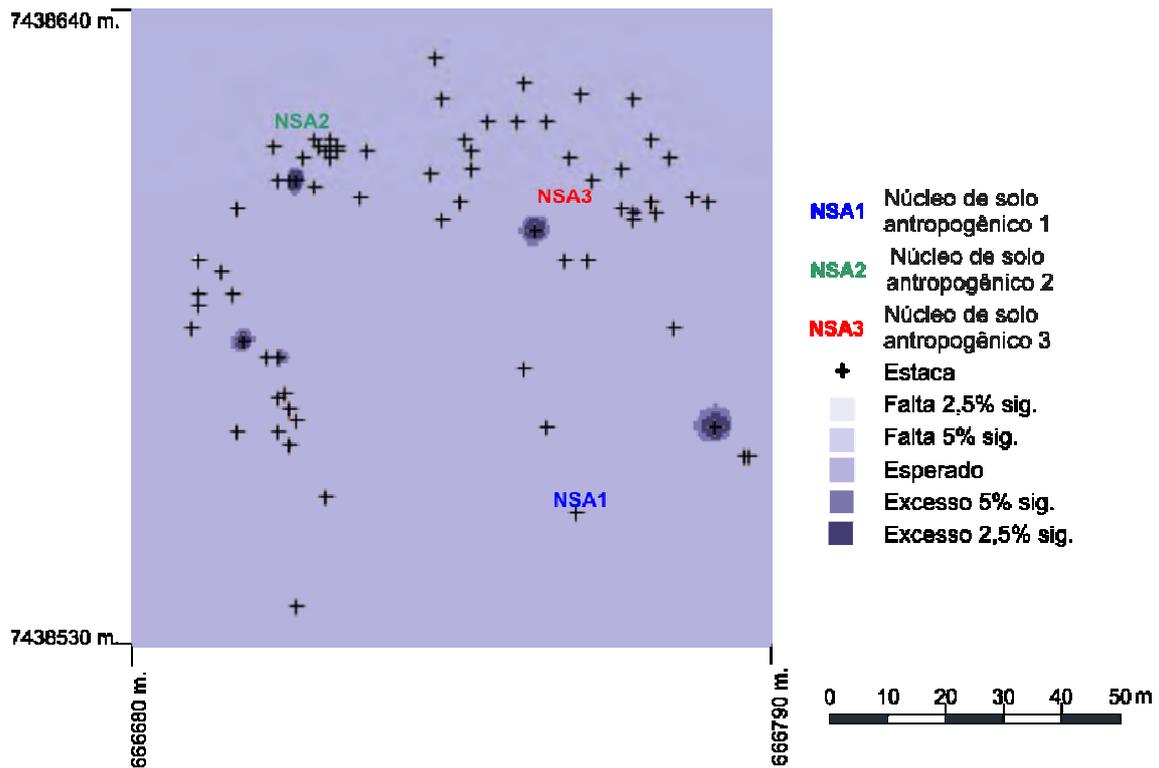


FIGURA 89 - Mapa temático do atributo tratamento superficial categoria 12 (TS_12).

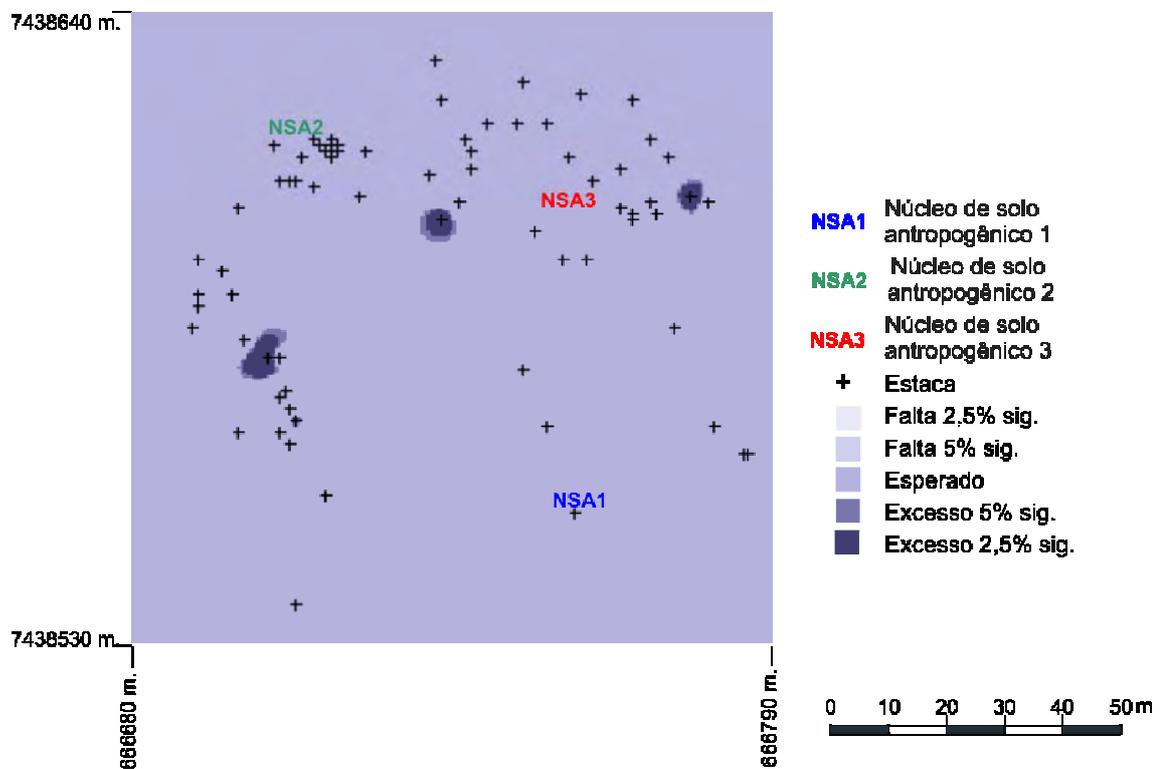


FIGURA 90 - Mapa temático do atributo tratamento superficial categoria 19 (TS_19).

Os mapas temáticos dos resíduos padronizados TS_6, TS_12 e TS_19 ilustram as figuras 88, 89 e 90, respectivamente, nos quais se observa pontos de concentração de TS_6 e TS_12 no entorno do NSA2. Os pontos de concentração do TS_19 encontram-se dispersos no Sítio Piracanjuba.

1.13.1.4 Tabela de contingência excepcionais

Dos 292 fragmentos cerâmicos que possuem ao menos um atributo_categoria considerado como excepcional, 167, ou seja, 57%, foram coletados no NSA1. Isto posto, optou-se por estudar separadamente esses fragmentos segundo o seu local de coleta. O estudo desenvolvido por meio de gráficos de coluna analisa primeiramente os fragmentos coletados no NSA1 e posteriormente os fragmentos restantes coletados em todas as outras estacas e NSAs.

Os gráficos de coluna dos atributos_categorias excepcionais coletados no NSA1, revela:

- (a) quanto à classe ilustrada na figura 92 – ausência de fragmentos das categorias 6 (parede e base), 15 (rolete de confecção), 16 (ombro), 20 (borda com furo de suspensão), 22 (polidor de sulco), sendo a categoria 21 (borda com suporte para tampa) a que apresenta o maior número de fragmentos (dez);
- (b) quanto ao tratamento superficial ilustrado na figura 92 – todas as categorias consideradas como excepcionais estão presentes, sendo a categoria 1 (sem tratamento interno/externo) e a 10 (lustro interno/alisamento externo) as que apresentam o maior número de fragmentos;

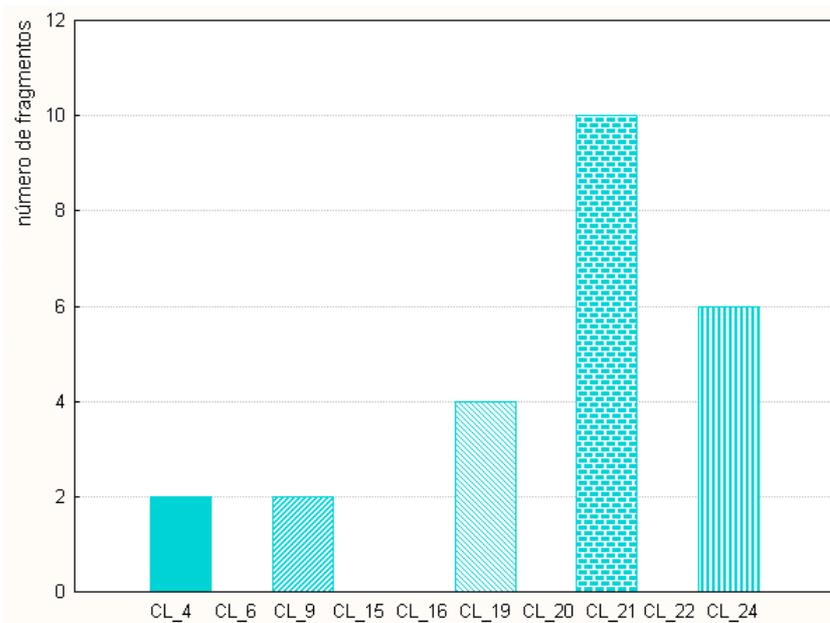


FIGURA 91 - Gráfico de colunas dos fragmentos cerâmicos coletados no NSA1 que apresentam o atributo classe em suas categorias excepcionais.

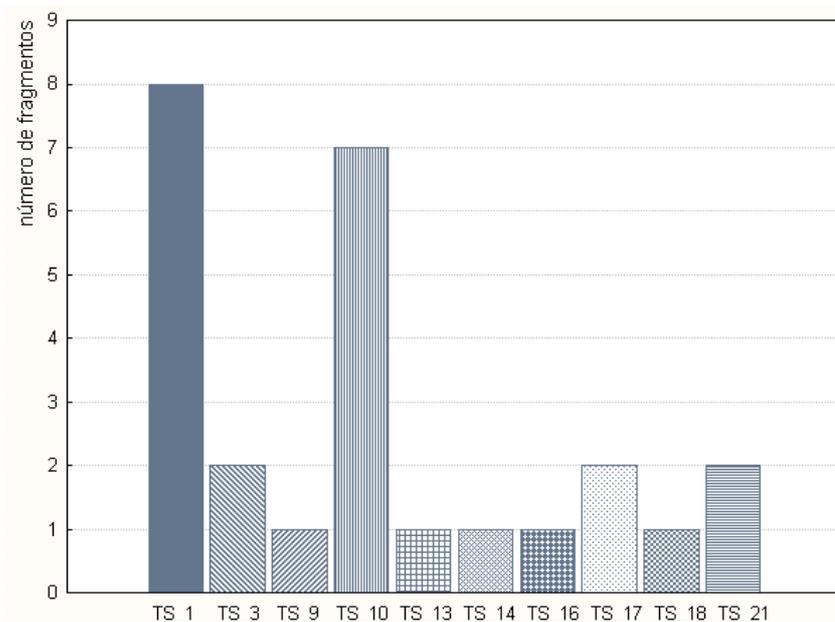


FIGURA 92 - Gráfico de colunas dos fragmentos cerâmicos coletados no NSA1 que apresentam o atributo tratamento superficial em suas categorias excepcionais.

(c) quanto à decoração interna ilustrada à figura 93 – ausência de fragmentos das categorias 5 (corrugado), 21 (engobo branco/inciso no contorno da garganta), 31

(engobo branco/laranja) e 32 (engobo vermelho pintado), sendo a categoria 38 (sem descrição) a que apresenta o maior número de fragmentos;

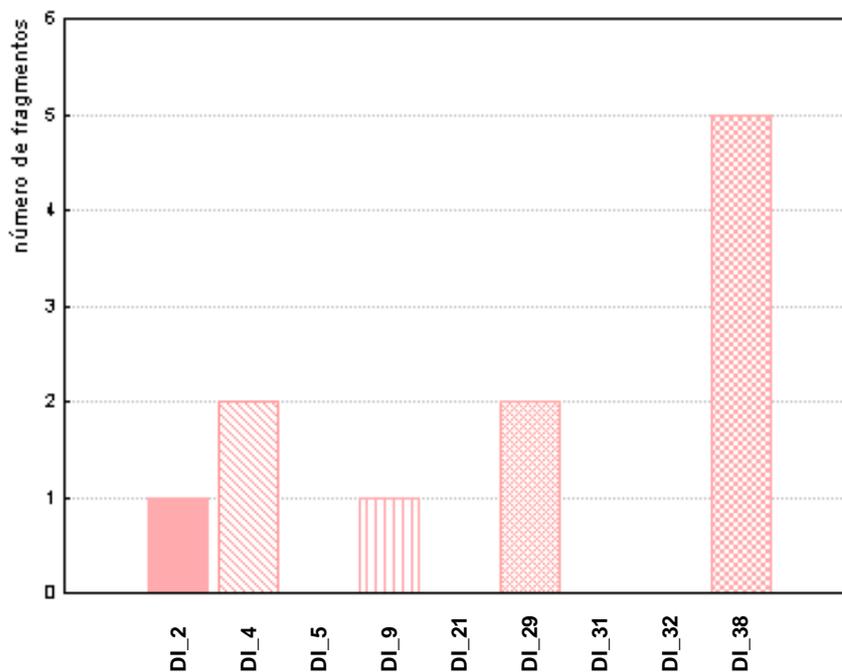


FIGURA 93 - Gráfico de colunas dos fragmentos cerâmicos coletados no NSA1 que apresentam o atributo decoração interna em suas categorias excepcionais.

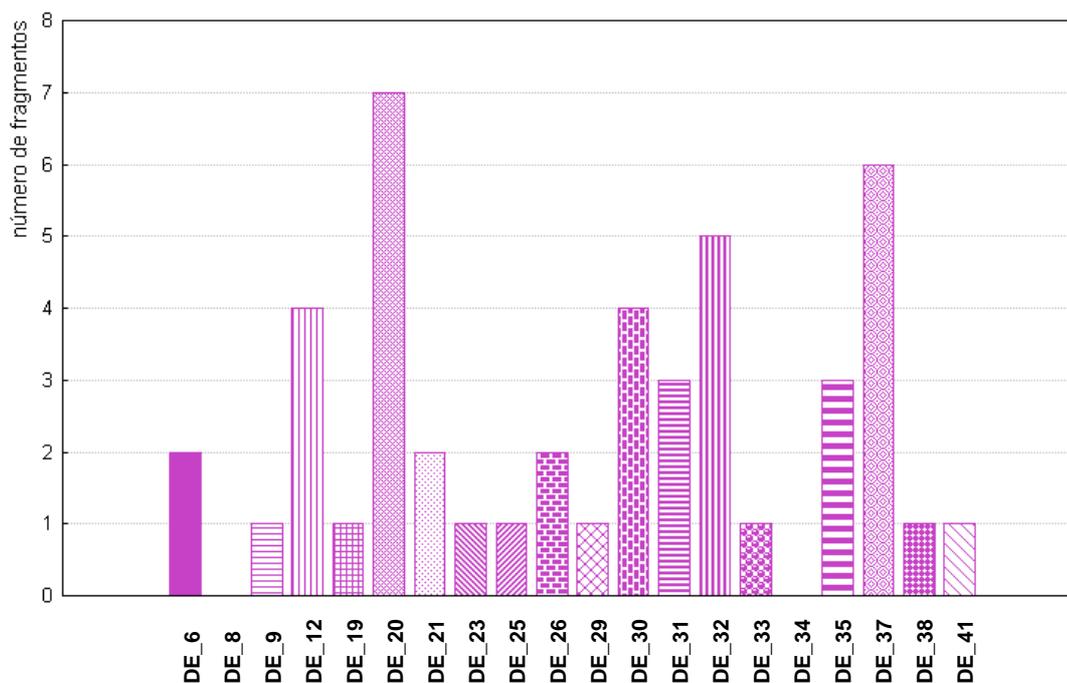


FIGURA 94 - Gráfico de colunas dos fragmentos cerâmicos coletados no NSA1 que apresentam o atributo decoração externa em suas categorias excepcionais.

- (d) quanto à decoração externa ilustrada na figura 94 – ausência das categorias 8 (pinçado) e a 34 (sem descrição), estando todas as outras categorias de decoração externa excepcionais presentes, com destaque para a categoria 20 (serrungulado) que apresenta o maior número de fragmentos;

O cruzamento da tabela de contingência excepcionais permitiu ainda que se detectassem os atributos_categorias exclusivos do NSA1, ou seja, que não ocorrem em nenhum outro fragmento coletado ao nível do solo no Sítio Arqueológico Piracanjuba.

Os atributos_categorias exclusivos bem como as suas distribuições de frequência encontram-se ilustrados na figura 95.

É conveniente destacar a categoria de antiplástico 19 (mineral + carvão + caco moído), que dentre os fragmentos cerâmicos coletados com atributos_categorias excepcionais apresenta o maior número ocorrências e é exclusivo do NSA1.

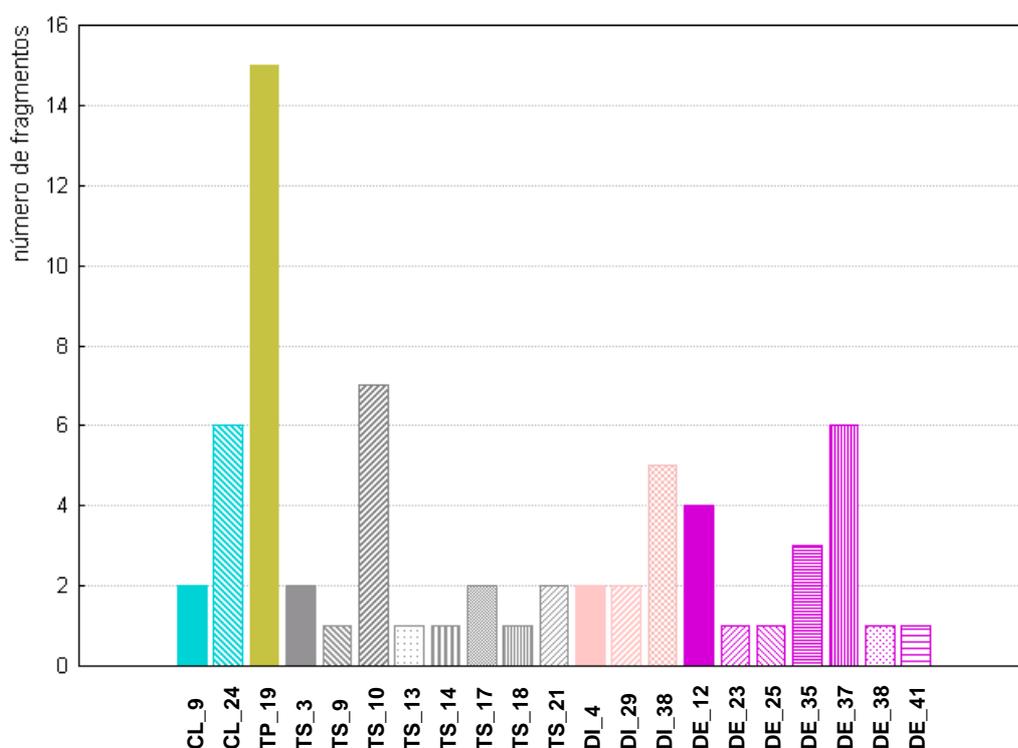


FIGURA 95 - Gráfico de colunas dos fragmentos cerâmicos de ocorrência exclusiva no NSA1 de atributos_categorias excepcionais

Dentre os fragmentos com atributos_categorias excepcionais coletados em locais diferenciados do Sítio Piracanjuba, à exceção do NSA1, pode-se dizer:

- (a) quanto à classe ilustrada à figura 96 – as categorias 6 (parede e base) e 16 (ombro) são exclusivas do NSA2, ou seja, os fragmentos com as características destes atributos_categorias foram coletados apenas no NSA2, e a categoria 20 (borda com furo de suspensão) é exclusiva da estaca 12. A categoria 21 (borda com suporte para tampa) apresenta a maior dispersão, bem como o maior número de fragmentos coletados;
- (b) quanto ao tratamento superficial – a diversidade das categorias de tratamento superficial encontra-se praticamente confinada ao NSA1, apenas a categoria 1 (sem tratamento interno/externo) possui 4 fragmentos coletados na estaca 63 e a categoria 16 (polimento externo/sem tratamento interno) 1 fragmento coletado no NSA2;

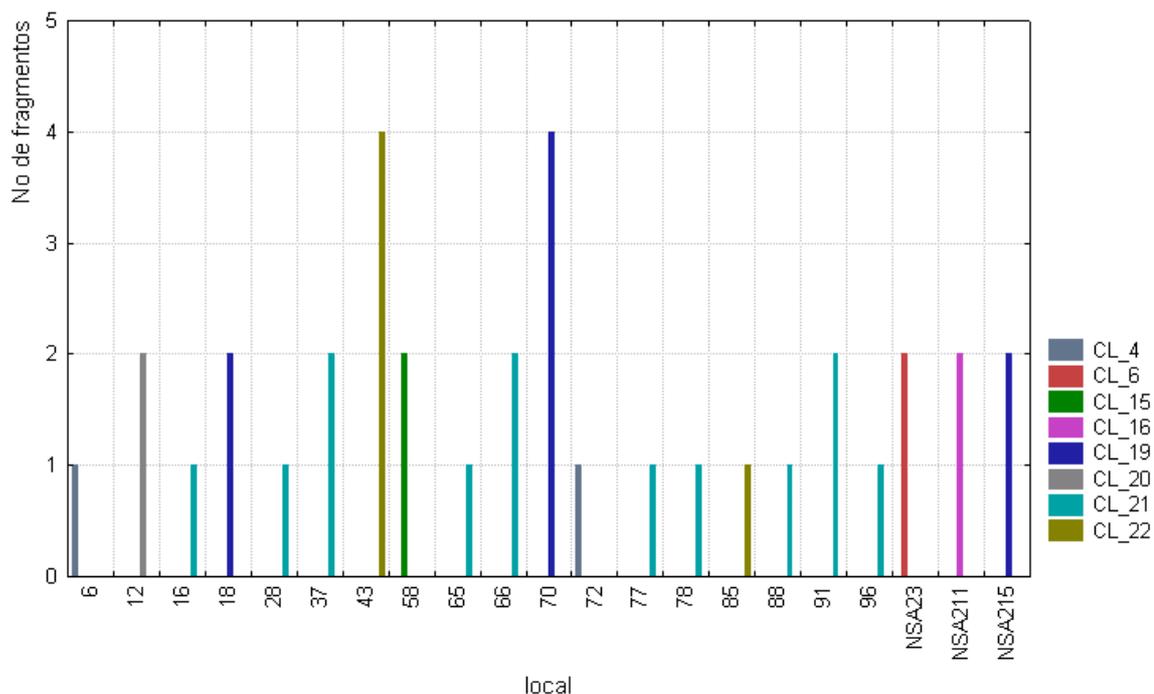


FIGURA 96 - Gráfico de colunas dos fragmentos cerâmicos coletados no Sítio Piracanjuba, exceto o NSA1, que apresentam o atributo classe em suas categorias excepcionais.

(c) quanto à decoração interna ilustrada na figura 97 – a categoria 5 (corrugado) é exclusiva da estaca 73 e a categoria 21 (engobo branco/inciso no contorno da garganta) bem como a 32 (engobo vermelho pintado), do NSA2. Os fragmentos com decoração interna 19 (canelado) apresentam maior frequência e dispersão.

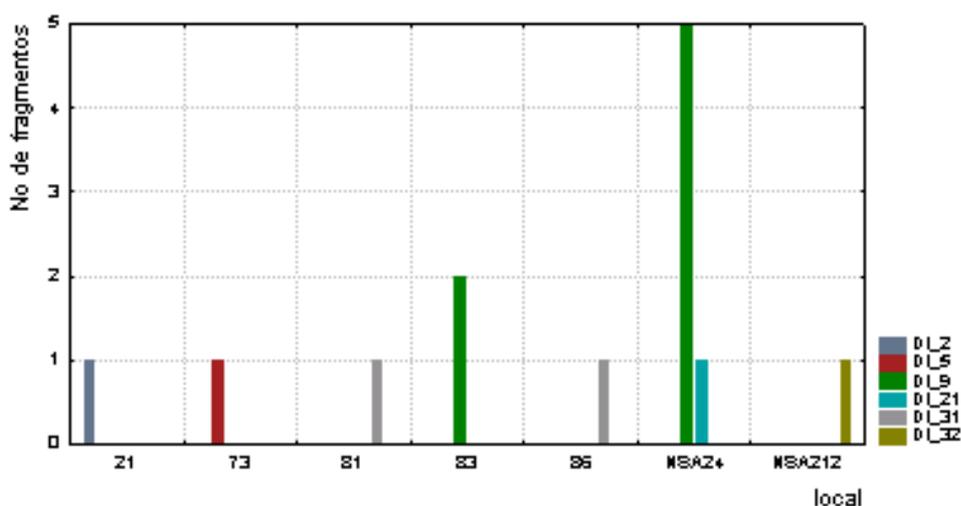


FIGURA 97 - Gráfico de colunas dos fragmentos cerâmicos coletados no Sítio Piracanjuba, exceto o NSA1, que apresentam o atributo decoração interna em suas categorias excepcionais.

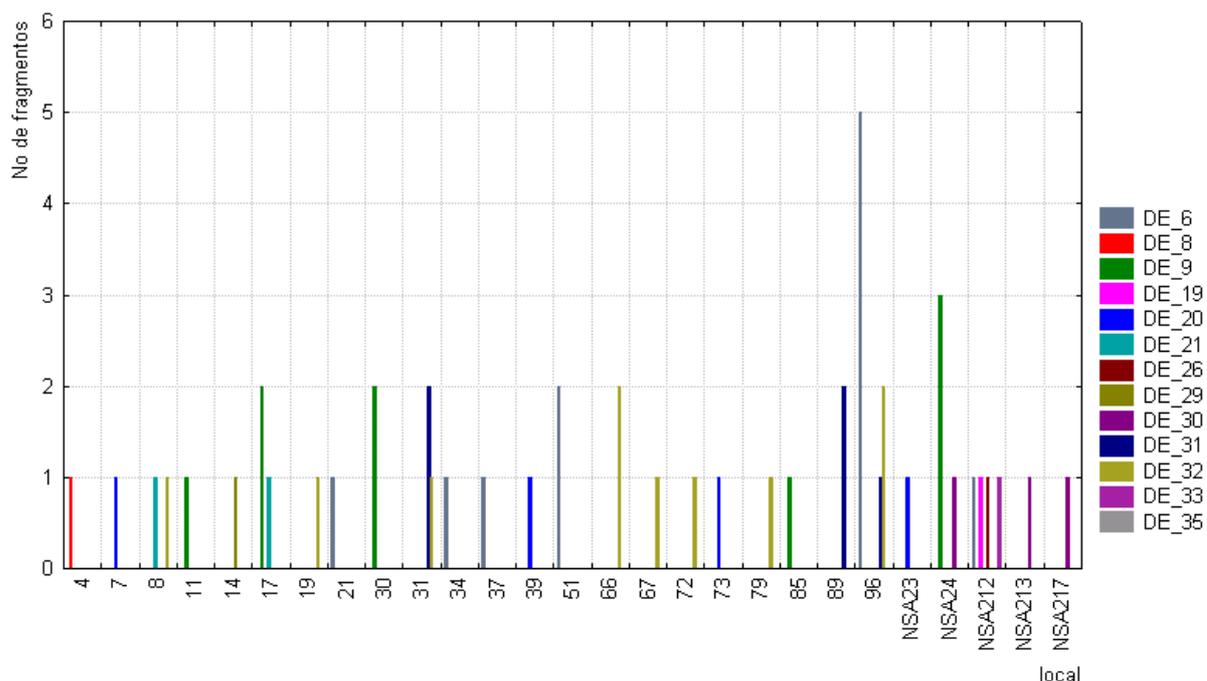


FIGURA 98 - Gráfico de colunas dos fragmentos cerâmicos coletados no Sítio Piracanjuba, exceto o NSA1, que apresentam o atributo decoração externa em suas categorias excepcionais.

- (d) quanto à decoração externa, ilustrada na figura 98, os fragmentos encontram-se bastante dispersos e não se detectou nenhuma categoria exclusiva, sendo as categorias 20 (serrungulado) e a 32 (engobo vermelho/pintado) as mais freqüentes.

1.14 CONCLUSÕES DA MINERAÇÃO DE DADOS

Quanto ao Sítio Piracanjuba pode-se dizer que dentre os fragmentos coletados, os de parede aparecem em maior número e distribuídos homogeneamente de forma proporcional em todo o sítio.

O antiplástico padrão do sítio é o do tipo 13 (material mineral + caco moído), onde as associações: material mineral de granulometria fina + caco moído de granulometria fina, material mineral de granulometria média + caco moído de granulometria grossa e material mineral de granulometria grossa + caco moído de granulometria média ocorrem mais frequentemente.

A técnica de manufatura roletado, o tratamento superficial alisado interna e externamente, a queima apresentando uma secção transversal com presença do núcleo central escuro e uma camada interna e externa clara, e a decoração interna e externa lisa, são preponderantes.

Quanto ao estado de conservação, 98% dos fragmentos coletados ao nível do solo apresentam bom estado de conservação.

O NSA2 e a maioria das estacas apresentam um perfil similar.

Os valores da inércia total obtidos na análise das tabelas fcomum - 0,08653, fraros - 0,88920 e fraríssimos - 2,2077, mostram que a dependência dos atributos_categorias às estacas e NSAs se estabelece na razão inversa do número de ocorrências do atributo_categoria, ou seja, quanto menor a ocorrência do atributo_categoria mais vinculado

ele se encontra em determinadas estacas e NSAs.

O estudo conjunto dos mapas temáticos dos resíduos padronizados dos atributos_categorias levou à delimitação de áreas que não obedecem ao padrão estabelecido para o Sítio Piracanjuba, caracterizadas em termos da concentração/falta de fragmentos cerâmicos de determinados atributos_categorias.

O NSA1 difere do restante do sítio quanto ao material constituinte e tratamento superficial dos fragmentos cerâmicos ali coletados.

Encontram-se presentes no NSA1 fragmentos cerâmicos de antiplástico do tipo 1 (material mineral), com ocorrência mais freqüente do que no restante do sítio, do tipo 19 (material mineral + caco moído + carvão), que ocorrem exclusivamente, e do tipo 13 (material mineral + caco moído), onde as diferenças estão ao nível da granulometria dos materiais, encontrando-se concentrados os materiais de granulometria média e grossa, em detrimento dos de granulometria fina.

Todas as categorias de tratamento superficial dos fragmentos cerâmicos coletados no Sítio Piracanjuba têm sua representação no NSA1, sendo que algumas encontram-se bastante concentradas, polimento interno/alisamento externo, polimento interno/externo e a categoria 11 que não apresenta descrição, são de ocorrência exclusiva, e limitada a poucos fragmentos o alisamento interno sem alisamento externo, lustro externo/alisamento interno, lustro interno/alisamento externo, brunidura/polimento interno, polimento interno/sem tratamento externo e as categorias 17, 18 e 21 sem descrição.

Quanto à decoração o NSA1 apresenta concentração de fragmentos cerâmicos decorados externamente com engobo vermelho e decoração externa 36, sem descrição, e um número menor do que o esperado para o padrão do sítio de fragmentos com decoração externa de engobo branco, pintado, entalhado e corrugado. Observa-se ainda baixa incidência de fragmentos de parede angular.

O NSA3 e seu entorno podem ser caracterizados pela diversidade de categorias de

atributos de fabricação que ali se concentram.

Observa-se a concentração de fragmentos de espessura média grossa e grossa, bem como os de decoração interna de engobo vermelho, engobo laranja e pintado. Quanto à decoração externa é grande a diversidade de categorias que ocorrem mais frequentemente do que o esperado para o padrão do sítio (apresentam concentração), são elas: entalhado, inciso, corrugado, escovado, engobo vermelho, engobo branco, engobo vermelho/branco, engobo laranja e as categorias 31 e 32 (sem descrição).

O atributo queima que aparece com grande homogeneidade em suas categorias 3, 4, 5 e 6 no sítio, apresenta no NSA3 e entorno a concentração de fragmentos com queima da categoria 1, seção transversal sem presença de núcleos, com cor uniforme, variando do laranja tijolo ao amarelo, e categoria 2, seção transversal, sem presença de núcleos com cor uniforme variando do cinza-claro ao pardo.

É conveniente ressaltar, ainda, a concentração de fragmentos cerâmicos constituídos de material mineral, antiplástico da categoria 1, e a concentração de caco moído de granulometria média na composição do antiplástico do tipo 13, em detrimento do caco moído de granulometria fina e grossa.

A detecção de fragmentos de base modelados à mão suscitou dúvidas na Prof^a Dra Neide Faccio Barrocá que voltou a campo e reavaliou os mesmos. Concluiu que apesar desses fragmentos estarem registrados na planilha eletrônica de dados fornecida para análise com esses atributos, cuja associação foi detectada pela análise de correspondência, os mesmos não se confirmam. Pode-se então dizer que a análise feita, aliada ao conhecimento do perito da área de arqueologia, colaborou na retificação dos dados dos atributos registrados para esses fragmentos cerâmicos.

Convém citar a ocorrência de fragmentos cerâmicos de parede angular com decoração externa de engobo vermelho/branco com pontos de concentração em comum na área delimitada pelas coordenadas 666680 - 666790 E e 7438587 - 7438640 N.

Com o intuito de melhor analisar espacialmente as áreas diferenciadas pelos atributos_categorias construiu-se o mapa ilustrado na figura 99, onde o critério de separação das áreas não obedece nenhum parâmetro formal, apenas a análise visual dos mapas temáticos dos resíduos padronizados dos atributos_categorias e a análise da mineração de dados.

Apesar do NSA1 apresentar o maior número de fragmentos coletados, esses fragmentos encontram-se mais condensados, ocupando uma área proporcionalmente menor do que as estacas similares ao NSA2 e ao NSA3, que se apresentam bastante dispersas. O NSA3 e seu entorno apresentam a mesma dispersão observada nas estacas e no NSA2.

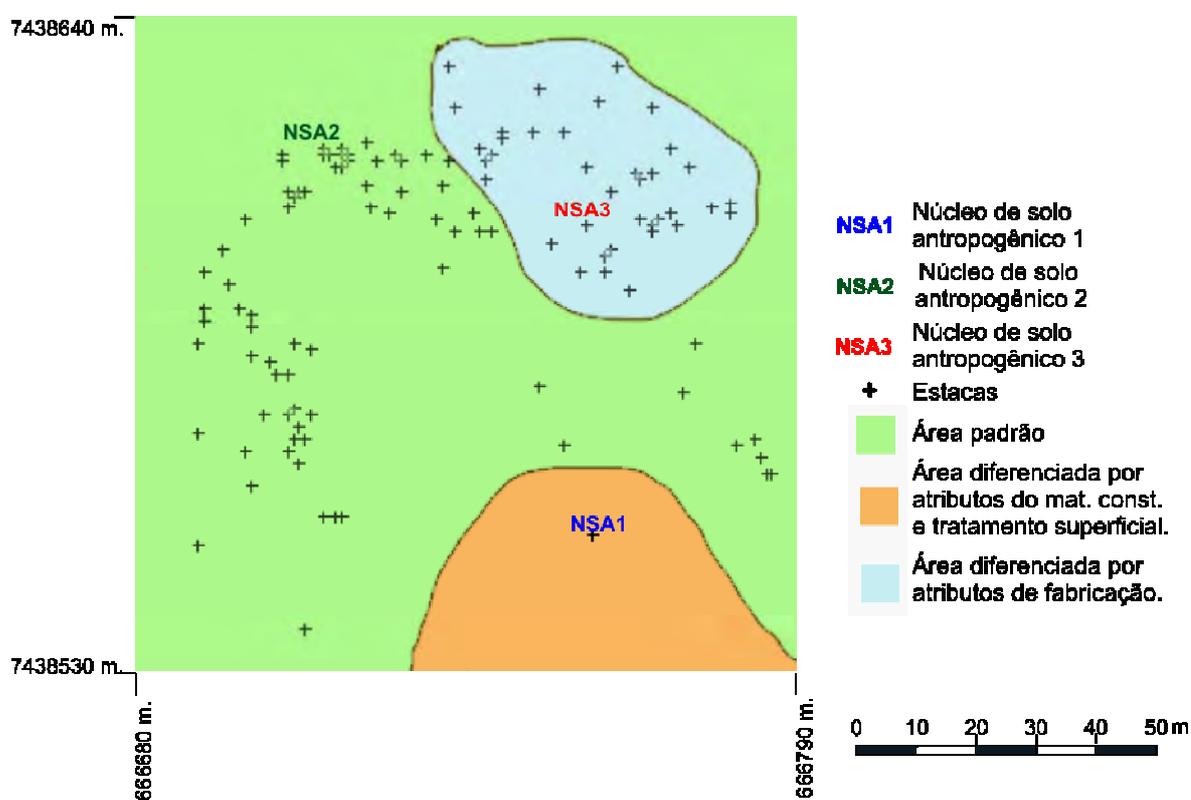


FIGURA 99 - Mapa temático de áreas diferenciadas quanto aos atributos_categorias do Sítio Piracanjuba.

CONCLUSÃO

A técnica da análise de correspondência mostrou-se hábil em explicitar os relacionamentos/associações intra-estacas e NSAs e intra-atributos_categorias, bem como inter estacas e NSAs e atributos_categorias dos fragmentos cerâmicos selecionados para a análise do Sítio Piracanjuba

A natureza e a localização desses relacionamentos/associações apresentados sob a forma de mapas temáticos dos atributos_categorias construídos pela espacialização dos resíduos padronizados em modelos numéricos de terreno tornam fácil e rápida a tarefa de extração de informações, sejam elas relativas a um específico atributo_categoria ou a um conjunto de atributos_categorias.

Os resultados obtidos da análise de correspondência aliados ao conjunto de mapas temáticos forneceram amplo arsenal de estudo, permitindo aos peritos da área arqueológica levantar inúmeras questões a partir das análises desenvolvidas.

Podem ser citadas a título de exemplo:

- o comportamento do material constituinte que se apresenta de maneira diferenciada no NSA1 pode corroborar a sua extemporaneidade?
- As diferentes composições preferenciais do antiplástico da categoria 13 encontram explicação no esgotamento de jazidas minerais, na manifestação de preferências das oleiras no objetivo de ajustar as técnicas ao tipo de função que desejavam dar às vasilhas?
- qual a hipótese capaz de justificar a diversidade de categorias de tratamento superficial no NSA1?
- qual a hipótese capaz de justificar a diversidade de categorias de decoração externa e interna detectada no NSA3 e entorno?

- as áreas delimitadas por similaridade entre estacas e NSAs, apresentam padrões distintos, uma se apresenta condensada em torno de 1 NSA e as outras duas apresentam grande dispersão de estacas no entorno dos NSAs. Pode-se ver aí indícios de ocupação habitacional diferenciada?

Raciocínios similares com hipóteses a serem testadas podem ser desenvolvidos tendo por base os resultados encontrados e o conhecimento *a priori* dos peritos da área de arqueologia.

Portanto, conclui-se que a metodologia da descoberta de conhecimento indireta, bem como a técnica estatística da análise de correspondência aliada à representação espacial dos resíduos padronizados, mostrou-se eficaz em fornecer uma visão sistêmica do Sítio Arqueológico Piracanjuba, capaz de contribuir para o conhecimento dos grupos indígenas que outrora habitaram o local.

6 REFERÊNCIAS

- ANDEBERG, M. R. **Cluster Analysis for Applications**. New York. Academic Press, 1973.
- AURÉLIO, M. et al. **Descoberta de Conhecimento e Mineração de Dados**. Apostila, ICA - PUC - Rio de Janeiro, 1999.
- BARTHOLOMEW, D. J. et al. **The Analysis and Interpretation of Multivariate Data for Social Scientists**. Florida. Chapman & Hall, 2002.
- BAXTER, M. J. et al. **Compositional data analysis in archaeometry**. Presented at CoDAWork, Girona, Spain, 2003. Disponível em <http://ima.udg.es/Activitats/CoDaWork03/paper_baxter_Beardah1.pdf>. Acesso em junho de 2004.
- BERRY, M.; LINOFF, G. **Data Mining Techniques: For marketing, sales and customer support**. Wiley Computer Publishing, USA, 1997.
- BIGOLIN, N.M. **Data Mining: Conceitos e Técnicas**, Escola Regional de Informática (ERI 2000), pp.233-250, Foz de Iguaçu, Tubarão, 2000.
- CAMPOS, I. O. **Análise de Técnicas Para o Estabelecimento de Modelos Digitais de Terreno – DTM**. Dissertação de Mestrado, Escola de Engenharia de São Carlos – Universidade de São Paulo, São Paulo, 1994.
- CHAIB FILHO, H. **Aplicações de Análise Fatorial de Correspondências para Análise de Dados**. In: Notas em Matemática Aplicada; SBMAC. São Carlos, SP, 2004.
- CRATOCHVIL, A. **Data Minig Techiques in supporting Decision Making**. Dissertação de Mestrado, 1999. Disponível em <<http://www.nici.ru.ni/~idak/teaching/leiden/mastertheses.html>>. Acesso agosto 2004.
- FACCIO, N. B. **O Estudo do Sítio Arqueológico Alvim no Contexto do Projeto Paranapanema**. Dissertação de Mestrado. USP, 1992.
- FACCIO, N. B. **Arqueologia dos cenários das ocupações horticultoras da Capivara, Baixo Paranapanema**. Tese de Doutorado em Arqueologia - FFLCH, Universidade de São Paulo, São Paulo, 1998.
- FAQ – Frequently Asked Questions. **Mineração de Dados**.1996. Disponível em <<http://www.rpi.edu/faq.html>>. Acesso dezembro 2000.
- FAYYAD, U.; PIATETSKY- SHAPIRO, G.; SMYTH, P. **From data mining to knowledge discovery: an overview**. In: FAYYAD, U.; PIATETSKY- SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R., Eds. *Advances in knowledge discovery*, p. 1- 36. Cambridge: MIT Press, 1996.
- FAYYAD, U.;PIATESKY-SHAPIRO, G.; SMYTH, P.**The KDD Process for Extracting Useful Knowledge from Volumes of Data**. Communications of the ACM, Vol.39, No. 11,

p.27-34. Nov. 1996.

FELGUEIRAS, A. C.; CÂMARA, G. **Modelagem numérica de terreno**. In: CÂMARA, G. Introdução à ciência da geoinformação. 2001. Disponível em <http://www.dpi.inpe.br/gilberto/livro/introd/cap7_mnt.pdf> . Acesso maio 2005.

FELGUEIRAS, A. C. **Modelagem Numérica de terreno**. In: CÂMARA, G.; MEDEIROS, J. S. Geoprocessamento para projetos ambientais. 1998. Disponível em <http://www.dpi.inpe.br/gilberto/tutoriais/gis_ambiente/4mnt.pdf>. Acesso maio 2005.

FERREIRA, O. C.; JÚNIOR, P. S. **Recursos audiovisuais para o ensino**. EPU Ltda, São Paulo, 1975.

FIGUERAS, S. M. e GARGALLO, P. **Análisis Exploratorio de Datos**. 5campus.com, Estadística. 2003. Disponível em <<http://www.5campus.com/leccion/aed>>. Acesso março de 2004.

FREITAS, A. A. **Understanding the crucial differences between classification and discovery of association rules: a position paper**. ACM SIGKDD Explorations Newsletter, v.2 n.1, p. 65-69, junho de 2000.

GREENACRE, M. J. **Theory and Applications of Correspondence Analysis**. Academic Press, London, 1984.

GROEBEL, M.; GRUENWALD, L. **A survey of data mining and knowledge discovery software tools**. SIGKDD Explorations, v. 1, p. 20- 33, 1999.

HABERMAN, S. J. **Analysis of Qualitative Data: Introductory Topics**. Academic Press, New York, 1978.

HAN, J. et al. **Spatial Clustering Methods in Data Mining: A Survey**. Anais School of Computing Science.Simon Fraser University. Canada, 2001.

HARRISON, T. H. **Intranet Data Warehouse**. Editora Berkeley Brasil, São Paulo, 1998.

IBM. **IBM's Data Mining Technology**. White Paper, 1996.

IPHAN – Instituto do Patrimônio Histórico e Artístico Nacional. Disponível em <<http://iphan.gov.br>>. Acesso em agosto de 2004.

KOPERSKI, K. et al. **Spatial Data Mining: Progress and Challenges Survey Paper**. SIGMOD'96 Workshop. on Research Issues on Data Mining and Knowledge Discovery (DMKD'96), Montreal, Canada, June 1996. Disponível em: <<http://db.cs.sfu.ca/sections/publication/kdd/kdd.html>>. Acesso outubro 2002.

KOPERSKI, K; HAN, J.; ADHIKARY, J. **Mining Knowledge in Geographical Data**, accepted by IEEE Comuter , 1998. Disponível em <ftp://ftp.faz.sfu.ca/pub/cs/han/pdf/geo_survey98.pdf>. Acesso março 2003.

LANDIM, P. M. B. **Introdução aos métodos de estimação espacial para confecção de mapas** . DGA, IGCE, UNESP/Rio Claro, Lab. Geomatemática, Texto Didático 02, 2000. Disponível em <<http://www.rc.unesp.br/igce/aplicada/textodi.html>>. Acesso abril 2005.

LANDIM, P. M. B. **Geologia Quantitativa: Introdução à análise estatística de dados geológicos multivariados**. Departamento de Geologia Aplicada. Instituto de Geociências e Ciências Exatas. UNESP / Campus de Rio Claro, 2001.

LANDIM, P. M. B., MONTEIRO, R. C. & CORSI, A. C. **Introdução à confecção de mapas pelo software SURFER**. DGA, IGCE, UNESP/Rio Claro, Lab. Geomatématica, Texto Didático 08, 2002. Disponível em <<http://www.rc.unesp.br/igce/aplicada/textodi.html>>. Acesso abril 2005.

LOPES, P. R.; FACCIO, N. B. **Análise da Cerâmica Guarani do Sítio Arqueológico Lagoa Seca II**. 2001 Disponível em <http://www.cibergeo.org/agbnacional/VICBG-2004/Eixo5/e5%20219.htm#_ednref1>. Acesso setembro 2004

MENDONÇA NETO, M. G. **Mineração de Dados**. Anais da 6a Escola Regional de Informática da Regional 2 do Estado de São Paulo, São Carlos, São Paulo, 2001.

MENDONÇA NETO, M. G., ALMEIDA, M. **O Uso de Interfaces Abundantes em Informação para Exploração Visual de Dados** In: Brazilian Workshop on Human Factors in Computer Systems , v.1. p.256 – 268, Florianópolis, 2001.

MICHELOUD, F. X. **Correspondence Analysis**. Pully, French, 1997. Disponível em <<http://www.micheloud.com/FXM/COR/E/>>. Acesso em setembro de 2006.

MOBERG, C. **Introdução à Arqueologia**. Edições 70, Lisboa, 1986.

MORAIS, J. L.; FACCIO, N. B. **O trabalho do Barro no Sítio Arqueológico Piracanjuba - Piraju, São Paulo**. Relatório de Pesquisa do ProjPar. USP, São Paulo, 2004.

MOXON, B. **Defining Data Mining** .DBMS Data Warehouse Supplement, August 1996. Disponível em <<http://www.dbmsmag.com/9608d53.html>> . Acesso em setembro 2003.

MUNITA, C. S. et al. **Characterization of Brazilian pre-historic ceramics by neutron activation analysis**. Progress Report 2000 - 2001 Research Reactor Center . Disponível em <<http://www.ipen.br/biblioteca/progress/pdfs/CRPQ/>> Acesso em julho 2004.

MUNITA, C. S. et al. **Chemical study of some ceramics from Brazilian Northeast**. Journal of Radioanalytical and Nuclear Chemistry, 2004, vol. 259, iss. 2, pp. 305-309(5) Kluwer Academic Publishers. Disponível em <<http://www.ingenta.com/isis/browsing/TOC/ingenta?issue=pubinfobike://klu/jrnc/2004/00000259/00000002>>. Acesso em julho 2004.

NAMIKAWA, L. M. et al. **Modelagem Numérica de Terreno e Aplicações**. INPE, São José dos Campos, 2003. Disponível em <<http://mtc-m12.sid.inpe.br/col/sid.inpe.br/marciana/2003/03.10.11.36/doc/publicacao.pdf>>. Acesso em maio de 2006.

NAGPAUL, P. S. **Guide to Advanced Data Analysis using IDAMS Software**. Nova Delhi, India. Disponível em <<http://www.unesco.org/webworld/idams/advguide/TOC.htm>>. Acesso em setembro de 2006.

NEVES DO VALE, M. **Agrupamentos de Dados: Avaliação de Métodos e Desenvolvimento de Aplicativo para Análise de Grupos**. Dissertação de Mestrado,

Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2006. Disponível em <http://www.maxwell.lambda.ele.puc-rio.br/cgi-bin/db2www/PRG_0487.D2W/INPUT?CdLinPrg=pt>. Acesso em julho de 2006

PEREIRA, J. C. R. **Análise de Dados Qualitativos: Estratégias Metodológicas para as Ciências da Saúde, Humanas e Sociais.** 3a. ed., EDUSP, São Paulo, 2004

PIATETSKY-SHAPIRO, G. **Knowledge discovery in real databases:** A report on the IJCAI- 89 Workshop. AI Magazine, Vol. 11, No. 5, p. 68- 70, Special issue, Jan. 1991.

ROBRAHN GONZÁLEZ, E. M. **A ocupação ceramista pré-colonial do Brasil Central.** Tese de Doutorado em Arqueologia - FFLCH, Universidade de São Paulo, São Paulo. 1996

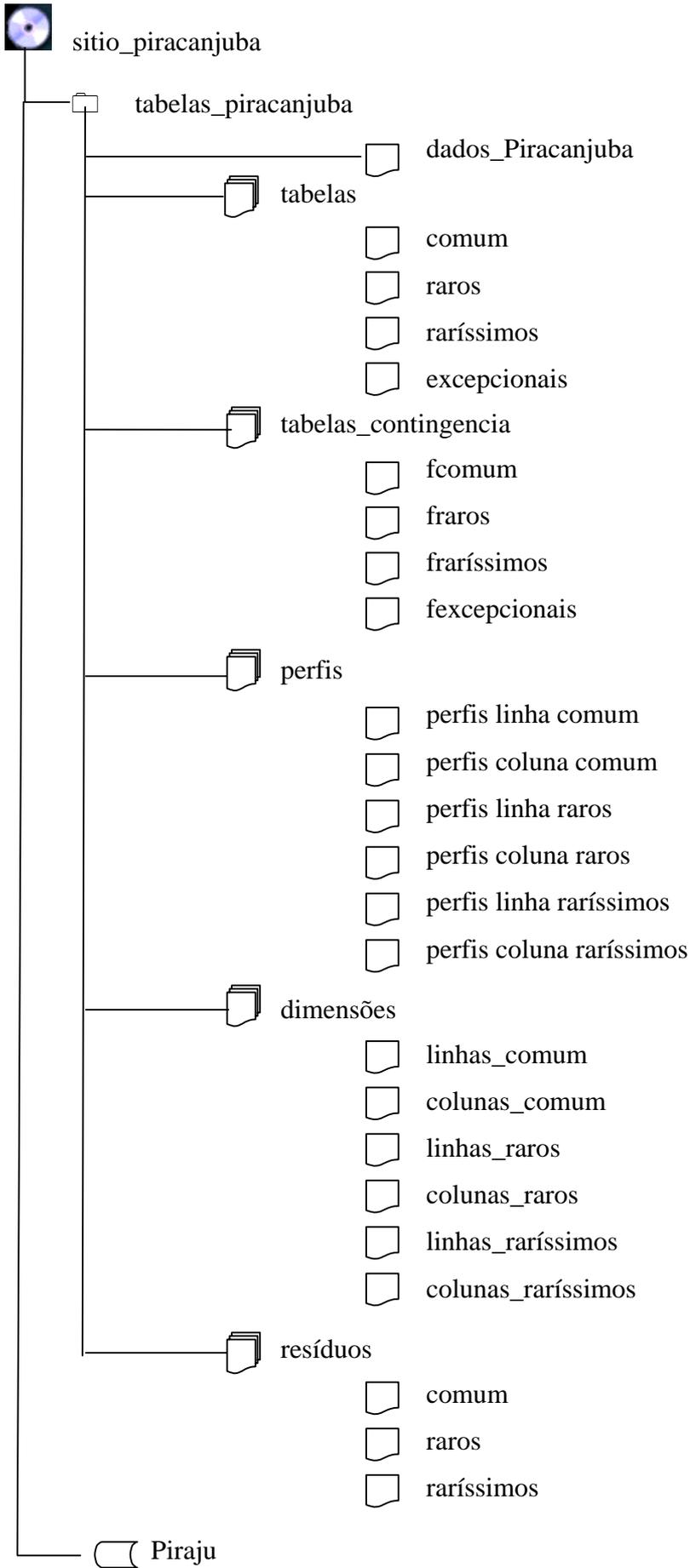
SANTOS, A. J. F. et al. **Uso de interfaces abundantes em informação para mineração visual de dados.** Relatório técnico RT-NUPERC-2000-4/p. (Relatório técnico do estado da arte), Universidade de Salvador, Salvador. 2000

SILVA, R.L. **Desenvolvimento de um modelo matemático para geração de superfícies por interpolação linear a partir do gradiente do terreno.** Anais do Congresso Brasileiro de Cadastro Técnico Multifinalitário - COBRAC 2004. Florianópolis, 2004. Disponível em <http://www.geodesia.ufsc.br/Geodesia-online/arquivo/cobrac_2004/147.pdf>

VON WANGENHEIM, A. **Técnicas Estatísticas para Reconhecimento de Padrões.** Disponível em < <http://www.inf.ufsc.br/%7Eawangenh/RP/estatisticas.html> > .Acesso setembro de 2004.

WICHERN, D. W., JOHNSON, R. A., **Applied Multivariate Statistical Analysis,** Prentice Hall, 5a. ed., 2001

APÊNDICE A – Mídia Digital



sendo:

 dados_Piracanjuba planilha eletrônica contendo os registros dos fragmentos cerâmicos, fornecida pela equipe do ProjPar

 tabelas

 comum planilha eletrônica contendo os registros dos fragmentos cerâmicos com atributos_categorias considerados comuns

 raros planilha eletrônica contendo os registros dos fragmentos cerâmicos com atributos_categorias considerados raros

 raríssimos planilha eletrônica contendo os registros dos fragmentos cerâmicos com atributos_categorias considerados raríssimos

 tabelas_contingencia

 fcomum planilha eletrônica contendo a distribuição de frequência da tabela comum

 fraros planilha eletrônica contendo a distribuição de frequência da tabela raros

 raríssimos planilha eletrônica contendo a distribuição de frequência da tabela raríssimos

 perfis

 perfis linha comum planilha eletrônica contendo os perfis linha, o perfil marginal das linhas expressos em porcentagem e o gráfico de linhas dos perfis linha e do perfil marginal das linhas da tabela fcomum.

 perfis coluna comum planilha eletrônica contendo os perfis coluna, o perfil marginal das colunas expressos em porcentagem e o gráfico de linhas dos perfis coluna e do perfil marginal das colunas da tabela fcomum.

 perfis linha raros planilha eletrônica contendo os perfis linha, o perfil marginal das linhas expressos em porcentagem e o gráfico de linhas dos perfis

linha e do perfil marginal das linhas da tabela fraros.

 perfis coluna comum planilha eletrônica contendo os perfis coluna, o perfil marginal das colunas expressos em porcentagem e o gráfico de linhas dos perfis coluna e do perfil marginal das colunas da tabela fraros.

 perfis linha raríssimos planilha eletrônica contendo os perfis linha, o perfil marginal das linhas expressos em porcentagem e o gráfico de linhas dos perfis linha e do perfil marginal das linhas da tabela raríssimos.

 perfis coluna raríssimos planilha eletrônica contendo os perfis coluna, o perfil marginal das colunas expressos em porcentagem e o gráfico de linhas dos perfis coluna e do perfil marginal das colunas da tabela raríssimos.

 dimensões

 linha_comum planilha eletrônica contendo as coordenadas das derivadas principais, a massa, a qualidade da representação, a contribuição a inércia total, a contribuição total e relativa para a inércia das linhas da tabela fcomum.

 coluna_comum planilha eletrônica contendo as coordenadas das derivadas principais, a massa, a qualidade da representação, a contribuição a inércia total, a contribuição total e relativa para a inércia das colunas da tabela fcomum.

 linha_raros planilha eletrônica contendo as coordenadas das derivadas principais, a massa, a qualidade da representação, a contribuição a inércia total, a contribuição total e relativa para a inércia das linhas da tabela fraros.

 coluna_raros planilha eletrônica contendo as coordenadas das derivadas principais, a massa, a qualidade da representação, a contribuição a inércia total, a contribuição total e relativa para a inércia das colunas da tabela fraros.

 linha_raríssimos planilha eletrônica contendo as coordenadas das derivadas principais, a massa, a qualidade da representação, a contribuição a inércia total, a contribuição total e relativa para a inércia das linhas da tabela fraríssimos.

 coluna_raríssimos planilha eletrônica contendo as coordenadas das derivadas principais, a massa, a qualidade da representação, a contribuição a inércia total, a contribuição total e relativa para a inércia das colunas da tabela fraríssimos.

 resíduos

 comum planilha eletrônica contendo os resíduos padronizados da tabela fcomum.

 raros planilha eletrônica contendo os resíduos padronizados da tabela fraros.

 raríssimos planilha eletrônica contendo os resíduos padronizados da tabela fraríssimos.

 Piraju Banco de dados do tipo .spr, contendo a localização espacial das estacas e NSAs, valores dos resíduos padronizados, grade retangular regular, isolinhas e mapas temáticos dos atributos_categorias com falta/excesso de ocorrências.